Collecting text errors in a separate file

This solution is not for everyone. It's only for those who care how flawless the layout text you're working with is. Obviously, the layout designer and the editor should work together on the task of bringing the text to such a result. The problem is that there has not yet been a suitable tool to solve this problem. Now it's there, it's a script that runs in the InDesign.

This program allows you to collect in a separate text file all the words that the speller marks as erroneous. Such a file can be created immediately after placing the file in the book layout. And then the editor will work with these words. To change something, to leave something as it is. And then the layout designer will already make edits to the text. Such targeted work with words that require attention should improve the quality of text preparation.

Don't you care if there are grammatical errors in your layout?

It always upset me when there were a lot of editorial edits and mistakes. You can, of course, console yourself that the editor will subtract, and I will correct. But it's more like self-deception. It's your personal time to make changes.

And another such moment. Ideally, the text should be proofread in Word and corrected without errors. But the reality is that now the editor is only reviewing the text sent for layout. And the real cleaning is already done when the editor receives the layout. That's how it happened. This is not a bad thing, it's just that we need to take into account such a modern approach to organizing work with text.

If the designer has the knowledge and ambition to correct the text

Yes, InDesign has dynamic spell checking. But over a long period of time, excellent results can only be achieved by working in single-tasking mode. And we are primarily engaged in the layout. And mistakes, well, sometimes we fix mistakes, but in most cases these are actions based on the residual principle. And it's not that we're such dunces, but it's just that there's no convenient way in InDesign to enable viewing problematic words from a speller's point of view.

Turn it on so that it's the main task for a while.

The way it is now — the window **Check spelling (Edit > Spelling > Check spelling)** — this cannot be considered convenient. Well, indeed, InDesign shows the words one by one, which he considers erroneous, you press the button **Skip** to show the next "suspicious" word, and this process will very soon seem pointless: it is not known how many problematic words have been found, and the very approach of clicking on each word is time-consuming.

If the text is edited by the editor while reading the layout

It would be good to put all the problematic words together first, so the editor can view this list, and not catch the words in the text of the layout.

And the option of "putting all the problematic words together" is one of the possible ones. Russian words only, English words only, and those with a mixture of Russian and Latin letters, for example, can be arranged in such a way that it will definitely simplify the work of the editor in putting the text in order.

And of course, it would be useful to indicate on which pages such a word occurs in the text, this may be necessary at the editing stage.

Surely such dreams "it would be nice..." have occurred to many. Yes, there was no convenient tool to solve this problem. But now there is such a program. The layout designer can assemble into a text file all the words of the document that the speller considers to be erroneous.

Then it is advisable to organize the work with the file as follows:
1) the first text layout, just putting the text in order and assigning the necessary styles; getting a list of words that need to be checked.
2) the editor checks this list and the layout designer makes corrections to the layout.
3) full-fledged layout, which will be carefully proofread by the editor. With this approach, the probability of missing errors in the layout text will be incomparably less than with the usual version.

About careless mistakes in the text

Before viewing the errors, you need to make sure that the text has the correct indication of the selected language. It happens that there are Russian letters on the screen, but all the text is marked that it is English. O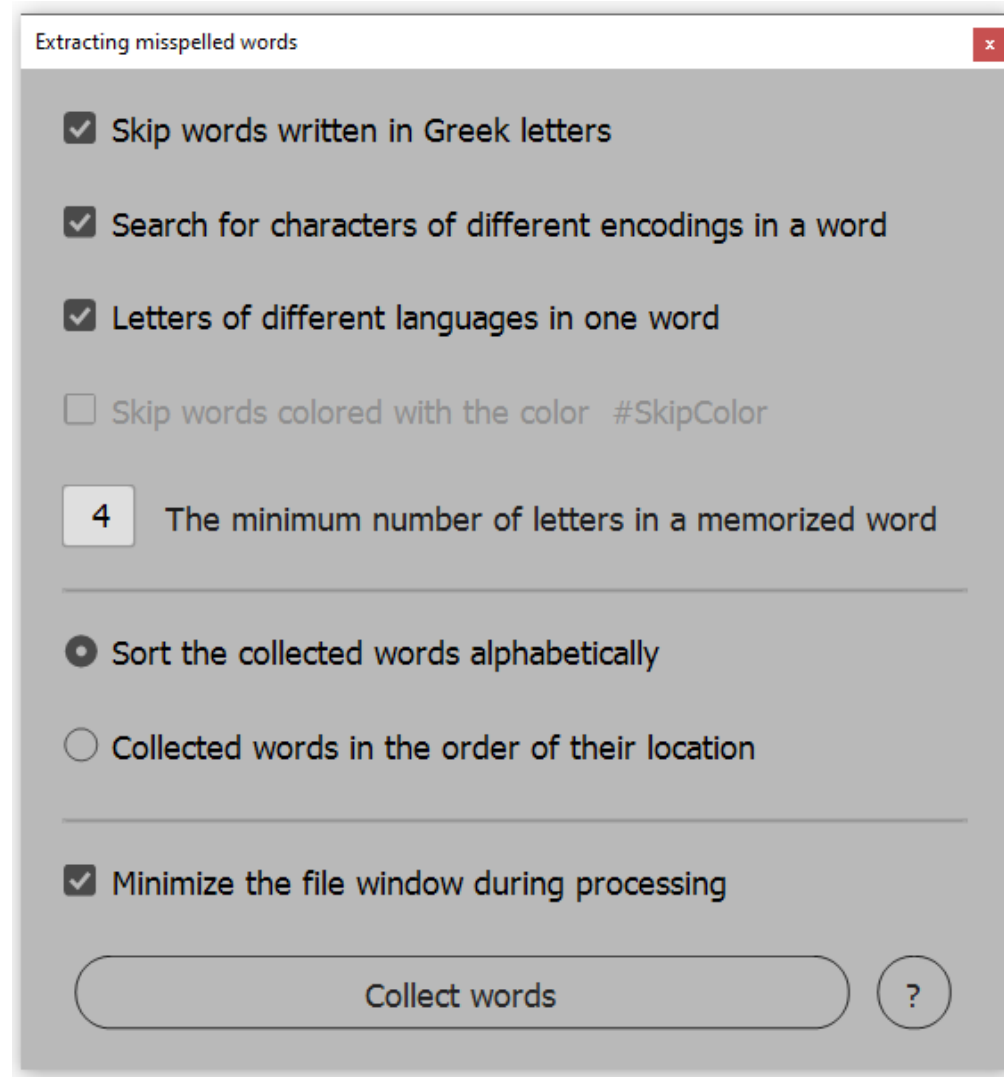r the English text has a language mark that it is Russian. The **Set Language By charCode.jsx** script has been created for the correct language markup of the text. It has his own manual. With its help, the preparation of texts for layout should become better.

Putting all the problematic words together

This word collection program is called **CollectSpellWords.jsx**. Its window is on the right.

Skip words written in Greek letters — do not include the symbols in the file, although the speller highlights them as errors;
Search for characters of different encodings in a word — This is sometimes found in Russian texts: letters similar in spelling turn out

---

Extracting misspelled words ☒

☑ Skip words written in Greek letters

☑ Search for characters of different encodings in a word

☑ Letters of different languages in one word

☐ Skip words colored with the color #SkipColor

[4]   The minimum number of letters in a memorized word

◉ Sort the collected words alphabetically

○ Collected words in the order of their location

☑ Minimize the file window during processing

(Collect words)   (?)

---

to be English. At the same time, the word is marked that its language is Russian. This can be for both lowercase and uppercase letters. In this program, these words are put together.
Letters of different languages in one word. This is like the case mentioned in the paragraph **About careless errors in the text**.
Skip words colored with the color #SkipColor — you can't fore-

see everything, and it may be that it doesn't make sense to send the words of some part of the layout to a file with errors. We need to make a service color **#SkipColor** and mark this area with it. If there is no such color in the layout, then the checkbox is unavailable.

The minimum number of letters in a memorized word — It hardly makes sense to put abbreviations and articles in the file, so you can set a threshold for the length of words. Anything less than that will be skipped. If there is **1**, then all the words found by the speller will be collected.

Sort... — two radio buttons that define how the collected words will be presented.

Minimize the file window during processing — when this checkbox is unchecked, we see the program passing through the text. This movement causes the screen to be constantly redrawn, and theoretically processing should go faster if you minimize the layout window.

When you run the script, the window **Check spelling** will appear in addition to this window. It should be positioned near the edge of the screen so that it does not overlap the progress bar window, which will be placed in the center of the screen.

The idea of a progress bar

It is unknown how many words the speller considers to be erroneous. And processing can take minutes, or even tens of minutes. It all de-

pends on the number of words that the speller has marked. And you can't have nothing happening on the screen all the time. We need to be sure 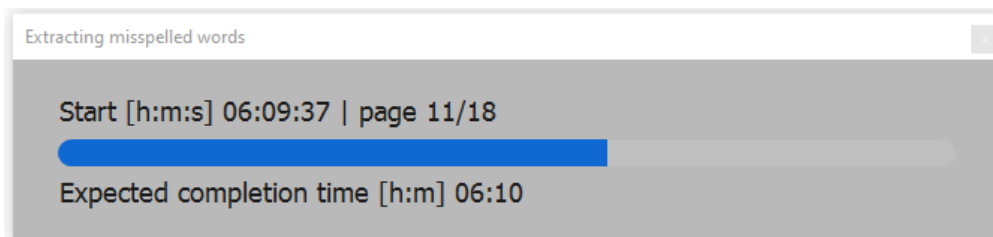that the program is not frozen and processing continues. Here's how it's done in the absence of information about the number of problematic words. It is known how many pages there are in the document; for each selected word, you can find out which page it is on; and you can make a progress bar for going through the pages of the document, rather than selecting the words found.

The start time of processing is displayed, the number of processed pages, and the one from which the words are currently being extracted, as well as the total number of pages. Here, the "total number of pages" is not only the pages of the document, but also all the master pages, because they may also contain errors.
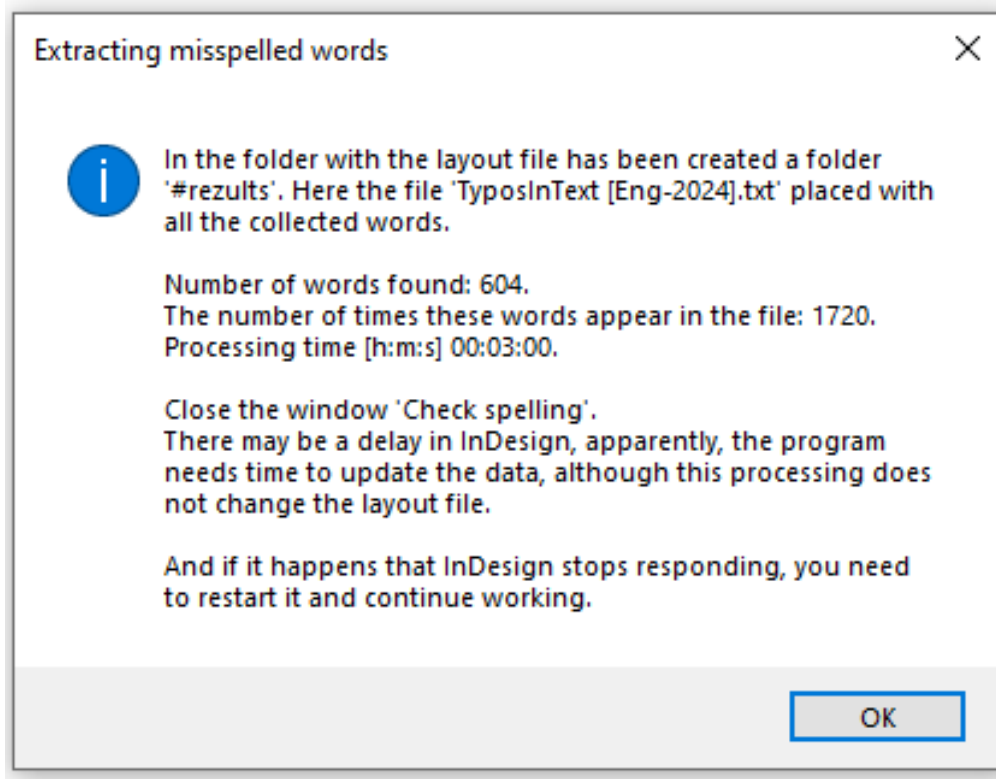
After processing the 10th page, the expected processing completion time is calculated and displayed in the progress bar window. Since the reality is that the processing speed slows down, this time is specified after processing each page.

The result of the work

The **#results** folder is created in the folder of the processed file, it will contain a txt file, the name of which matches the name of the indd file (UTF-8 encoding) with all the words found by the speller. There will be a message about this, as shown on the next page. This 18-page document was processed in 3 minutes, and it contains 604 erroneous words.

Extracting misspelled words

Start [h:m:s] 06:09:37 | page 11/18

Expected completion time [h:m] 06:10

The window **Check spelling** must be closed by yourself. There is an option to open a window in the scripting toolkit, but there is no way to close it.

Extracting misspelled words

> In the folder with the layout file has been created a folder '#rezults'. Here the file 'TyposInText [Eng-2024].txt' placed with all the collected words.
>
> Number of words found: 604.
> The number of times these words appear in the file: 1720.
> Processing time [h:m:s] 00:03:00.
>
> Close the window 'Check spelling'.
> There may be a delay in InDesign, apparently, the program needs time to update the data, although this processing does not change the layout file.
>
> And if it happens that InDesign stops responding, you need to restart it and continue working.
>
> **OK**

Simplest_Test.indd

Number of words found: 11.
The number of times these words appear in the file: 12.
Processing time [h:m:s] 00:00:03

== Misspelled words ==
Kiplimg : B
Tyutchev : A, −
alowance : 2
dream : 2
forse : 2
fountainhead : 1
iner : 1
krystal : 1
trulh : 2
untrue : 1

== Different encodings in one word ==
feel : 1

Is anyone ready to click the **Skip** button so many times to view all the words? Unlikely. And the editor or proofreader will prefer to work with the usual list of problematic words and make a decision on each of them, and for this these words must be extracted from the layout. Now there is such an opportunity.

The simple demo file
In folder Info there is two page file Simplest_Test.idml. There are errors in text, on master pages, on pasteboard, and the lines of explanations for these errors marked in blue. On the right is a script message about processing this file. It is the file Simplest_Test.txt in the folder **#rezults**.

After the word is a colon, and then the page where it occurs. If there are more than one page, they will be separated by commas. In this example, two words have a letter instead of a number. These are the misspelled words posted on the master page.
If the frame with the wrong word is on the desktop, then there will be a dash instead of the page number.
The words marked by #SkipColor not placed in the list of incorrect words.

How to use this file

The text should be placed in the Word editor. And then, when reading, mark the really problematic words with color. Maybe there will be an understanding that it would be good to put some of these words in a user dictionary. Color labeling is the easiest way to tell the typesetter which words need to be corrected.

Connecting user dictionaries

Probably, this guide should have started with this topic. But this program appeared later: the topic of the need to be able to exclude certain words from the area of attention of the speller, for example, surnames, arose in correspondence.

And since such a request has appeared, why not make a program. It's not difficult to assemble words from a file — each word in its own line. There is only one restriction on this dictionary word file — it should not contain footnotes. The script **WordsForDictionary.jsx** is for solve this task. It has his own manual.

Mikhail Ivanyushin
https://shop.dotextok.ru/en/
dotextok@gmail.com