

ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

УДК 004.915

Н. Б. Каледина, К. Т. Тарасевич

Белорусский государственный технологический университет

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТОВОГО ДОКУМЕНТА, ИМЕЮЩЕГО СИСТЕМАТИЧЕСКИЕ ОШИБКИ КОМПЬЮТЕРНОГО НАБОРА

В статье рассматриваются расширенные возможности процессора Microsoft Word, которые используются для правки текстового документа, в частности для быстрого исправления наиболее распространенных ошибок, возникающих при несоблюдении правил набора и верстки.

С целью выявления наиболее эффективного способа редактирования текста проведен ряд экспериментов, в процессе которых несколькими способами производилась правка текстового документа, содержащего большое количество ошибок набора. В ходе первого эксперимента правка текста осуществлялась вручную. Второе испытание предполагало редактирование документа с помощью стандартного инструмента поиска и замены. Далее правка текста производилась разработанными макросами, каждый из которых рассчитан на исправление определенного типа ошибки. В процессе четвертого эксперимента записанные ранее макросы были объединены в один, способный решить все поставленные задачи.

Разработана серия регулярных выражений, используемых для редактирования текста с помощью инструмента поиска и замены; на их основе созданы макросы, работающие с определенными видами ошибок, описаны особенности их функционирования; доказана эффективность использования единственного макроса, предусматривающего исправление всех видов ошибок. Оговорено, что макросы являются наиболее приемлемым, но не идеальным способом правки текста вследствие периодических сбоев в работе системы.

Ключевые слова: правка текста, текстовый процессор, правила набора и верстки, инструмент поиска и замены, макрос.

N. B. Kaledina, K. T. Tarasevich

Belarusian State Technological University

AUTOMATION OF PROCESSING OF THE TEXT DOCUMENT HAVING SYSTEMATIC ERRORS OF A COMPUTER SET

The article deals with advanced means of Microsoft Word, used for editing a text document, in particular, for quick correction of the most common errors that can occur because of inobservance of the rules of set and make-up.

In order to identify the most effective way to edit the text, a number of experiments were carried out. In the course of experiments editing a text with a lot of type errors was performed. In the course of the first experiment editing the text has been realized manually. The second test has involved editing by a standard search and replace tool. Next, editing the document has been made by devised macros, each of is supposed to correct of a specific type of errors. During the fourth experiment previously recorded macros has been combined into the one that can solve all problems.

A series of regular expressions used for editing a text by the find and replace tool has been worked out; macros working with certain kinds of errors, have been created, peculiarities of their functioning have been described; the efficiency of using the only macro, that provides correction for all types of errors, has been proved. It has been mentioned that macros are the most acceptable in the process of editing but they are often imperfect because of constantly system failures.

Key words: editing a text, word processor, the rules of set and make-up, the find and replace tool, macro.

Введение. Как известно, текст должен соответствовать правилам орфографии и пунктуации, а также нормам и традициям книжного набора. К последним относятся правила оформления заголовков, переносов слов; применение

знаков препинания, тире и дефисов; оформление кавычек и т. д. Однако зачастую рукописи содержат большое количество ошибок набора. Вот некоторые, на наш взгляд, наиболее распространенные из них:

1) наличие нескольких пробелов между словами, пробелов перед знаками препинания;

2) установка лишних разрывов строк и страниц, пустых абзацев;

3) обычный пробел после однобуквенных сокращений, например т. Иванов, г. Минск (необходимо, чтобы после них был установлен так называемый неразрывный пробел: т. Иванов, г. Минск);

4) присутствие пробела в начале строки;

5) установка обычного вместо необходимого неразрывного пробела между числами и наименованиями, например, 5 мм;

6) отсутствие неразрывных пробелов в стандартных сокращениях (и т. д.), между инициалами, инициалами и фамилией (С. И. Иванов), после символов №, §;

7) отбивка кавычек и скобок от заключенных в них слов, от знаков препинания;

8) использование английских машинописных или немецких кавычек вместо французских;

9) использование точки вместо запятой в десятичных дробях;

10) неправильное употребление различных видов тире [1, с. 41–44].

Зачастую перед редактором стоит необходимость осуществить качественную правку текста в сжатые сроки. Для решения поставленной задачи широко используется текстовый процессор Word, который по праву является на сегодняшний день одной из наиболее популярных программ для работы с документами. Программа позволяет создавать и редактировать документы любой сложности, при этом возможно выполнение большого количества различных операций с текстовой и графической информацией. Чаще всего рядовой пользователь не владеет приемами работы с автоматизированными средствами. Однако знание особенностей функционирования программы и некоторых хитростей позволит значительно ускорить правку текста и качественно улучшить процесс редактирования.

Целью данной работы послужило исследование эффективности редактирования текста следующими способами: 1) вручную; 2) с помощью инструмента поиска и замены; 3) путем создания нескольких макросов, каждый из которых исправляет определенный тип ошибки; 4) созданием одного макроса, решающего все проблемы. Для эксперимента был взят текст объемом 70 страниц, содержащий большое количество ошибок правил набора и верстки.

Основная часть. Наиболее часто в процессе редактирования текста пользователь прибегает к ручной правке. Согласно проведенному исследованию, данный способ работы с документом является проверенным, однако мало-

эффективным, особенно если речь идет о документах большого объема. Вследствие наличия таких, обусловленных человеческим фактором проблем, как пропуск ошибок из-за усталости или внешних раздражителей, заикливание на каком-либо одном конкретном виде ошибки, вычитку текста приходится осуществлять несколько раз, что значительно увеличивает время редактирования текста и снижает эффективность процесса.

Правку текста существенно упрощает мощный инструмент поиска и замены. Большинство пользователей прибегает к процедуре поиска для того, чтобы отыскать в документе конкретный фрагмент текста (слово, часть слова, выражение и т. д.). Однако расширенные возможности инструмента используются для создания регулярных выражений (систем литералов и подстановочных знаков, находящихся и заменяющих образцы текста), с помощью которых возможно некоторым образом автоматизировать правку текста, сделать процесс редактирования проще и быстрее. Так, например, знак «*» приемлем для поиска любой последовательности символов, знак «?» — одного символа; «<>» и «>» — для задачи определенного начала и конца слова. Для группировки подстановочных знаков и текста, а также для указания порядка обработки знаков следует использовать круглые скобки. Соответственно, выражение «к?t» позволяет найти слова «кот» и «кит»; «к*й» — «какой» и «который»; «<(сол)» — «солнце» и «солист», а «(соль)>» — «фасоль» [2].

В шаблонах поиска и замены необходимо учитывать следующие замечания.

В тексте вида «{1;}», символ точка с запятой (;) является настраиваемым. Данный символ задается в региональных настройках операционной системы, параметром «Разделитель элементов списка». Точка с запятой используется по умолчанию в русской версии Windows, в английской же принят символ запятая.

По одному и тому же шаблону с подстановочными знаками поиск вперед и назад может дать разные результаты.

Поиск с подстановочными знаками «@» и «{1;}» дает не одно и то же:

– например, при поиске вперед в тексте «111+222+333»: по шаблону «1@» будет найден один символ «1», а по «1{1;}» — цепочка символов «111»; по шаблону «{!2}@+» будет найден текст «111+», а по шаблону «{!2}{1;}+» ничего не будет найдено;

– при сложном поиске, особенно в обратном направлении, больше подходит «@». Например, поиск назад в тексте «Один Два Три» по шаблону «<[А-Я]{1;}[А-Я-я]{1;}>» не даст результата, в отличие от «<[А-Я]@[А-Я-я]@>».

Согласно приобретенному опыту более стабильным является «{1;}». Например лучше использовать «<[А-Я]{1;}>», а не «<[А-Я]@>». Результат одинаков, но первый вариант надежнее.

Подстановочный знак \n, где n=1,2,3...: можно использовать не только в поле «Заменить на» но и в поле «Найти». Например, если ищем по шаблону «(ма)\1», то будет найден текст «мама».

Замененный текст, в некоторых случаях, наследует формат впереди стоящего. Например, если искать «(?)X» и заменять на «\1Y», то, когда первый символ найденного текста подстрочный, а второй нет, после замены и

второй станет подстрочным. Поиск «X(?)» и замена на «Y\1» не приводит к наследованию.

Спецсимвол «!» при поиске с подстановочными знаками следует использовать осторожно. Например, поиск по шаблону «[!A]» не найдет графический объект, гиперссылку и т. п.

При поиске с подстановочными знаками лучше не применять знак конца абзаца «^13» вместе со скобками.

В ходе исследования при выполнении операций в окне НАЙТИ И ЗАМЕНИТЬ, а впоследствии — для записи макросов использовались различные комбинации подстановочных знаков, приведенные в таблице.

Примеры поиска и замены, используемые в процессе правки документа

Действие	Найти	Заменить
Замена разрыва строки на знак абзаца	^l	^p
Замена разрыва страницы на знак абзаца	^m	^p
Удаление лишних пробелов	^0032 {2;}	^0032
Удаление пробелов перед знаком абзаца	^0032^0013	^0013
Удаление лишних абзацев	^0013 {2;}	^0013
Удаление пробелов красной строки	^0013^0032(*)	^0013\1
Удаление пробелов перед знаками препинания	^0032([.,:;!/?])	\1
Неразрывный пробел после цифры перед буквой	([0-9]^0032([А-я])	\1^s2
Неразрывный пробел после однобуквенного слова	(<[А-я]>^0032([А-я])	\1^s2
Неразрывный пробел перед римскими цифрами	([А-я]^0032([IVXML])	\1^s2
Удаление лишних пробелов после раскрывающейся скобки	((
Удаление лишних пробелов перед закрывающейся скобкой))
Неразрывный пробел в инициалах	(<[А-Я]>).\(<[А-Я]>).\(<[А-Я]>)	\1.^s2.^s3
Неразрывный пробел после однобуквенных сокращений	(<[А-я]>).^0032([А-я])	\1.^s2
Неразрывный пробел после номера и знака параграфа	([№§])\([0-9])	\1^s2
Неразрывный пробел между цифрами	([0-9]^0032([0-9])	\1^s2
Замена дефисов и коротких тире на длинное тире в прямой речи в начале абзаца	^0013^0032-^0032	^0013—
	^0013^0032-	^0013—
	^0013-^0032	^0013—
	^0013^0032-^0032	^0013—
	^0013^0032-	^0013—
Замена английских машинописных и немецких кавычек на французские	"([А-я])	«\1
	([А-я])"	\1»
	ChrW(8220)·&·"([А-я])"	«\1
	"([А-я])"·&·ChrW(8221)	\1»
Замена дефисов и коротких тире на длинное тире в прямой речи	([.,:;!/?])^0032-^0032	\1—
	([.,:;!/?])^0032-	\1—
	([.,:;!/?])-^0032	\1—
	([.,:;!/?])^0032-^0032	\1—
	([.,:;!/?])-^0032	\1—
Замена дефиса перед цифрой на минус	([.,:;!/?])^0032-	\1—
	-([0-9])	^0150\1
	<[0-9]@).([0-9]@>)	\1,\2
	([0-9])\([0-9])	\1- \2
	Замена дефисов и коротких тире на длинное тире между словами	·
-		—
-		—
-		—
-		—

Данный способ правки текста, несмотря на его достоинства, содержит в себе следующие недостатки.

1. Необходимость запоминать огромное количество комбинаций подстановочных знаков.

2. Большой объем затрачиваемого времени.

3. Отдельные замены приходится делать несколько раз.

4. Так как в некоторых правилах существуют определенные нюансы и тонкости, произвести замену одним щелчком мыши не всегда представляется возможным. Как следствие, редактирование текста сводится к несколько автоматизированной, но, по своей сути, ручной правке.

5. Вероятность ошибок все так же велика.

6. Многие проблемы (такие, например, как расстановка надстрочных и подстрочных знаков) по-прежнему остаются нерешенными.

В ходе исследования было создано шесть макросов, каждый из которых решал конкретный тип задач.

Макрос «Первичное форматирование» выполняет следующие действия:

- 1) замена разрыва строки на знак абзаца;
- 2) замена разрыва страницы на знак абзаца;
- 3) удаление лишних пробелов;
- 4) удаление пробелов перед знаком абзаца;
- 5) удаление лишних абзацев;
- 6) удаление пробелов красной строки.

Макрос «Пробелы» предусматривает удаление лишних пробелов перед знаками препинания, внутри скобок и расстановку неразрывных пробелов после цифры перед буквой, после однобуквенного слова, номера и знака параграфа, перед римскими цифрами и между цифрами.

Макрос «Сокращения» используется для расстановки неразрывных пробелов после однобуквенных сокращений, между инициалами, а также в выражениях типа «и т. д.», «т. е.», «т. к.».

Макрос «Тире» решает задачи замены дефисов и коротких тире на длинное тире там, где это необходимо, а также замены дефисов на короткое тире в числовых диапазонах.

Макрос «Кавычки» заменяет английские машинописные и немецкие кавычки на французские.

Макрос «Специфические замены» включает в себя решение следующих задач:

- 1) замена точек на запятые в дробных числах;
- 2) замена дефиса перед цифрой на минус;
- 3) задачи, не решенные с помощью окна НАЙТИ И ЗАМЕНИТЬ:

- расстановка надстрочных знаков;
- расстановка подстрочных знаков.

Работа с макросами значительно повышает эффективность и точность правки текста, а также сокращает время редактирования документа. Разделение макросов на конкретные

группы является скорее преимуществом, нежели недостатком, поскольку некоторым образом индивидуализирует работу с документом (например, нет необходимости использовать макрос «Специфические замены» при редактировании художественных текстов, если заранее известно, что автор не использовал химические формулы, десятичные дроби и размерные обозначения типа «см²»). Однако следует соблюдать четкую последовательность действий: целесообразно начинать с макроса «Первичное форматирование», после чего переходить к использованию макроса «Пробелы» и только потом осуществлять остальные замены. В противном случае макросы придется запускать несколько раз. Также существенным недостатком являются периодические сбои в их функционировании, происходящие вследствие системных ошибок.

При записи последнего макроса, задачей которого стало исправление всех наиболее распространенных видов ошибок, использовались написанные ранее макросы, объединенные в единый код.

Полученные в ходе исследования данные позволяют построить сравнительную диаграмму, в которой отражена зависимость количества времени, затраченного на правку текста, от способа редактирования (рис. 1).

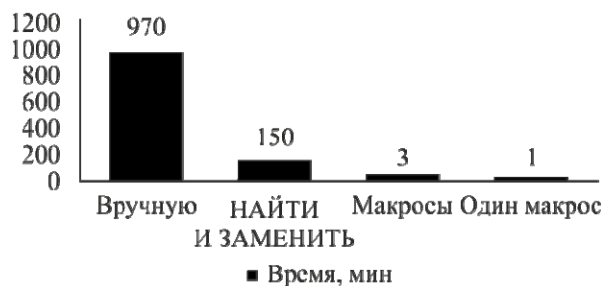


Рис. 1. Время, затраченное на правку текста

Сравнительная диаграмма, представленная на рис. 2, отображает процент ошибок, не исправленных в процессе правки текста каждым из способов.

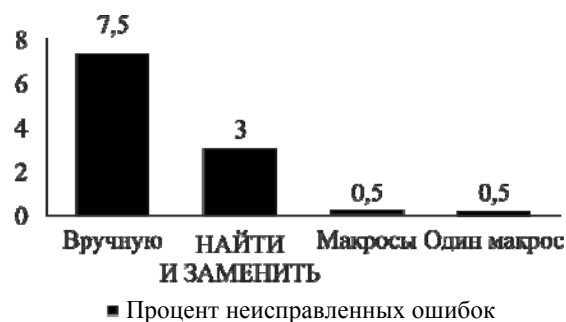


Рис. 2. Процент ошибок, не исправленных в процессе правки текста

Согласно проведенному исследованию, правка текста с помощью одного макроса является наиболее эффективным способом исправления ошибок, поскольку позволяет редактировать текст с высокой точностью в максимально короткий срок. Нет необходимости следить за конкретной последовательностью действий, поскольку она заранее прописана в программном коде. Однако, опять же, следует учитывать сбои в работе системы, случающиеся тем чаще, чем больше операций включает в себя макрос.

Заключение. Текстовый процессор Microsoft Word предоставляет широкие возможности

для работы с документами. Знание определенных приемов использования стандартных автоматизированных средств значительно упрощает процесс редактирования, делая его быстрее и эффективнее. Проведенное исследование показало, что наиболее приемлемым средством качественной правки текста являются макросы.

Однако их использование не гарантирует идеальный результат, поскольку периодически имеют место сбои и системные ошибки в их работе. Поэтому процесс редактирования документа должен включать в себя разумное сочетание ручной и автоматизированной работы.

Литература

1. Каледина Н. Б. Технология работы в текстовом процессоре Microsoft Word 2010. Минск: БГТУ, 2012. 278 с.
2. Поиск и замена текста с помощью регулярных выражений (расширенные возможности) // Microsoft Office. 2010. URL: <http://office.microsoft.com/ru-ru/word-help/HA102350661.aspx> (дата обращения: 26.02.2015).

References

1. Kaledina N. B. *Tekhnologiya raboty v tekstovom protsessore Microsoft Word 2010* [The technology of working in the text processor Microsoft Word 2010]. Minsk, BGTU Publ., 2012. 278 p.
2. *Poisk i zamena teksta s pomoshch'yu regulyarnykh vyrazheniy (rasshirennye vozmozhnosti)* [Find and replace text by using regular expressions (Advanced)]. Available at: <http://office.microsoft.com/ru-ru/word-help/HA102350661.aspx> (accessed 26.02.2015).

Информация об авторах

Каледина Наталья Борисовна — старший преподаватель кафедры полиграфических производств, Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: kaledina@belstu.by.

Тарасевич Ксения Тарасовна — студентка 2 курса кафедры редакционно-издательских технологий, Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь).

Information about the authors

Kaledina Natal'ya Borisovna — senior lecturer of the Department of Printing Technologies, Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: kaledina@belstu.by.

Tarasevich Kseniya Tarasovna — student of the Publishing and Printing Faculty, Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus).

Поступила 16.03.2015