

УДК 621.398
РАЗРАБОТКА ПОДСТРОКОВЫХ РЕГУЛЯРНЫХ ВЫРАЖЕНИЙ

Ю. В. ПАСЫНКОВА

Научный руководитель Э. И. ЯСЮКОВИЧ, канд. техн. наук, доц.
БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ

Регулярные выражения представляют собой формальный язык поиска и манипулирования подстроками текста – вставка, замена, удаление, основанный на использовании метасимволов. Работать с регулярными выражениями можно в различных программных приложениях, в том числе и в редакторе MS Word. Для этого достаточно на вкладке «Заменить» окна «Найти и заменить» в поле «Параметры поиска» отметить строку «Подстановочные знаки», прописать в текстовой строке «Найти» подстановочные знаки и, возможно, заполнить строку «Заменить на».

На сайте интернет [1], содержится конструктор регулярных выражений Javascript Regexp Evaluator [2], позволяющий в реальном времени составлять и проверять шаблоны регулярных выражений, который тестировался в таких браузерах, как Mozilla, IE 6.0 и Opera 7. Однако, поскольку в двух последних из них имеются ошибки в реализации Javascript, что может сказаться на правильной работе упомянутого конструктора.

Приведенные технологии работы с регулярными выражениями требуют большого объема знаний довольно сложной их грамматики. В связи с этим в настоящей работе была поставлена задача разработки приложения, позволяющего строить регулярные выражения для работы с текстом путем выбора элементов этих выражений в окне его интерфейса.

Разработанное приложения написано на языке C#, в котором используется класс Regexp пространства имен RegularExpression, содержащий методы: *IsMatch* – проверяющий наличие в тексте строки, соответствующей шаблону; *Match* – возвращающий первую подстроку, соответствующую шаблону; *Matches* – выполняющий поиск подстрок, соответствующих шаблону в виде коллекции типа *MatchCollection*; *Replace* – заменяющий все подстроки, соответствующие шаблону, новой строкой; *Split* – возвращающий массив строк, полученный разделением входной строки в местах соответствия шаблону [3].

Разработанное приложение, интерфейс которого приведен на рис. 1, позволяет по входному шаблону найти вхождения в файле форматов .txt или .doc и вывести результаты в форму richTextBox с отображением всего текста с выделенными фрагментами, соответствующими шаблону, или заменить подстроки, соответствующие шаблону, подстрокой, указанной в поле «Заменить на», которое появляется при выборе режима работы «заменить».



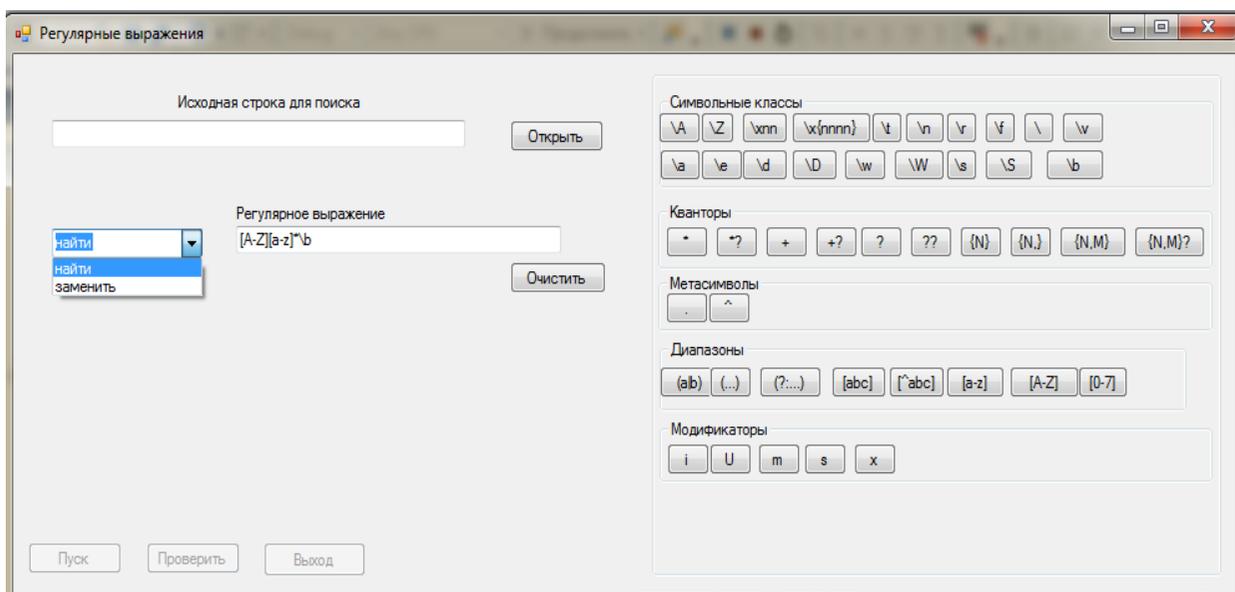


Рис. 1. Интерфейс приложения для работы с регулярными выражениями

Текстовое поле «Исходная строка для поиска» интерфейса предназначена для ввода исходной строки, в которой требуется произвести поиск. Кнопка «Открыть» используется для работы с текстовым файлом. Поле «Регулярное выражение» используется для ввода шаблона, которое может быть построено для режимов работы «найти» или «заменить». Для построения регулярного выражения можно использовать символьные классы, кванторы, метасимволы, диапазоны и модификаторы, приведенные в правой части окна интерфейса.

Запуск приложения на выполнения осуществляется кнопкой «Пуск». Кнопка «Проверить» предназначена для проверки шаблона, а «Выход» - для завершения работы приложения.

Разработанное приложение предусматривает два режима работы: с файлами или с текстовыми фрагментами.

СПИСОК ЛИТЕРАТУРЫ

1. <http://www.pcre.ru/eval>
2. **Гойвертс, Я.** Регулярные выражения. Сборник рецептов. Regular Expressions: Cookbook / Я. Гойвертс, С. Левитан. – СПб. : Символ-Плюс, 2010. – 608 с.
3. **Смит, Б.** Методы и алгоритмы вычислений на строках (regex) Computing Patterns in Strings / Б. Смит. – М. : Вильямс, 2006. – 496 с.