**Практическая работа № 1**

**Основы регулярных выражений**

**Цель работы:** изучение основ регулярных выражений на языке C#.

***Краткие теоретические сведения***

**1 Понятие регулярных выражений**

**Регулярное выражение** – это некий шаблон, составленный из символов и спецсимволов, который позволяет находить подстроки, соответствующие этому шаблону в других строках.

Регулярные выражения предоставляют такие возможности, как:

1) замена в строке всех одинаковых слов другим словом;

2) удаление в строке определенных слов;

3) выделение из строки необходимой части;

4) проверка соответствия строки заданному шаблону;

5) проверка, содержит ли строка заданную подстроку;

6) извлечение из строки все вхождения подстрок, соответствующие шаблону регулярного выражения.

**2 Класс Regex языка C#**

Для того, чтобы работать с регулярными выражениями необходимо подключить в начале программы пространство имен

*using System.Text.RegularExpressions;*

В C# работу с регулярными выражениями предоставляет класс *Regex*. Создание регулярного выражения имеет следующий вид:

*Regex myReg = new Regex([шаблон]);*

Здесь: *myReg* – имя созданного объекта;

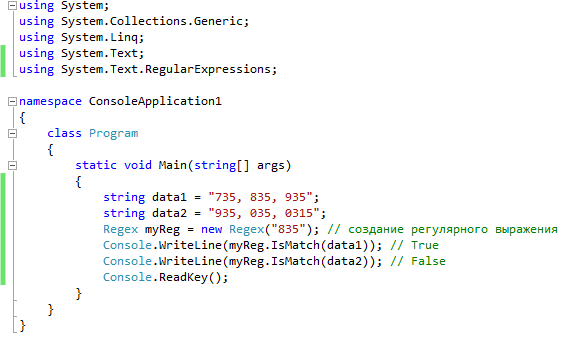
Конструктор вызывается при создании объекта (экземпляра) класса.

Здесь **[шаблон]** – это строка, содержащая символы и спецсимволы.

У *Regex* также есть и второй конструктор, который принимает дополнительный параметр – опции поиска.

**Задача 1.** Проверить вхождение шаблона в заданные выражения.

Точкой входа в консольное приложение является метод **Main()** главного класса программы.



Результат работы показан на рисунке 1.

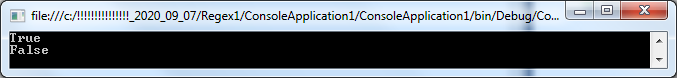


Рисунок 1 – Результат работы программы задачи 1

Здесь в качестве шаблона выступает однозначная строка "835". Далее использован метод **IsMatch()**, который проверяет, содержит ли заданная строка (data1, data2) подстроку соответствующую шаблону.

Таблица 1 – **Варианты заданий для задачи 1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № варианта | Шаблон |  | № варианта | Шаблон |
| 1 | Котов |  | 21 | Собакин |
| 2 | Кошкин |  | 22 | Сидоров |
| 3 | Овчаркин |  | 23 | Иванов |
| 4 | Тетерев |  | 24 | Барсуков |
| 5 | Аистов |  | 25 | Зоткин |
| 6 | Сусликов |  | 26 | Захаров |
| 7 | Хвостов |  | 27 | Зверев |
| 8 | Хомяков |  | 28 | Медведев |
| 9 | Стрижов |  | 29 | Зайцев |
| 10 | Коршунов |  | 30 | Волков |
| 11 | Воробьев |  | 31 | Псов |
| 12 | Кобылин |  | 32 | Овсов |
| 13 | Кроликов |  | 33 | Рыбин |
| 14 | Петухов |  | 34 | Тараканов |
| 15 | Уткин |  |  |  |
| 16 | Селезнев |  |  |  |
| 17 | Соловьев |  |  |  |
| 18 | Лисов |  |  |  |
| 19 | Тигров |  |  |  |
| 20 | Муравьев |  |  |  |

Рассмотрим методы для работы с регулярными выражениями.

**IsMatch()** проверяет, содержит ли строка хотя бы одну подстроку, соответствующую шаблону регулярного выражения. Работа этого метода показана в примере 1.

**Match()** возвращает первую подстроку, соответствующую шаблону, в виде объекта класса Match. Класс Match предоставляет различную информацию о подстроке – длину, индекс, само значение и другое.

**Задача 2.** Найти первую подстроку, соответствующую шаблону в заданном выражении.

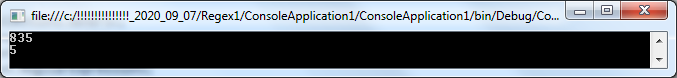
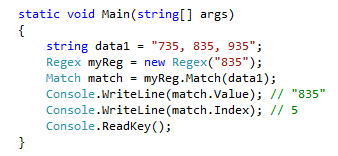


Рисунок 2 – Результат работы программы примера 2

Объект класса Match содержит поля:

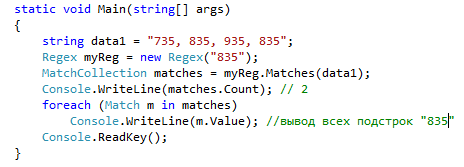
**Value** – найденное вхождение подстроки;

**Index** – начальная позиция вхождения (начиная с 0).

**Варианты заданий для задачи 2** взять из таблицы 1.

**Matches()** возвращает все подстроки, соответствующие шаблону, в виде коллекции класса *MatchCollection*. Каждый элемент этой коллекции соответствует классу *Match*.

**Задача 3.** Найти все подстроки, соответствующие шаблону в заданном выражении.



**foreach** – это цикл, выполняемый для всех элементов заданного объекта (например, массива). В данном примере осуществляется обращение ко всем объектам класса **Match** в объекте с именем **matches**.

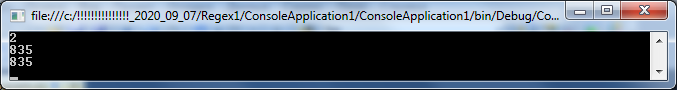


Рисунок 3 – Результат работы программы примера 3

**Варианты заданий для задачи 3** взять из таблицы 1.

**Replace()**возвращает строку, в которой заменены все подстроки, соответствующие шаблону, новой строкой.

**Задача 4.** Выполнить замену подстроки, соответствующей шаблону в заданном выражении.

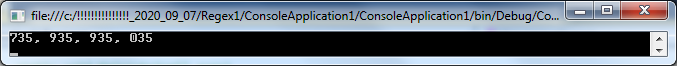
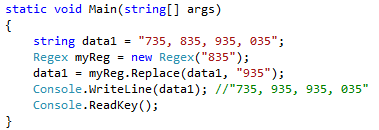


Рисунок 4 – Результат работы программы примера 4

**Варианты заданий для задачи 4** взять из таблицы 1.

**Split**() возвращает массив строк, полученный в результате разделения входящей строки в местах соответствия шаблону регулярного выражения.

**Задача 5.** Вывести массив строк, полученный в результате разделения входящей строки в местах соответствия шаблону регулярного выражения.

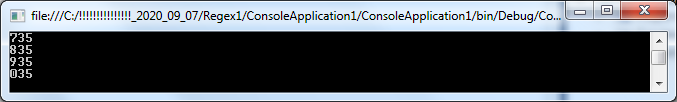
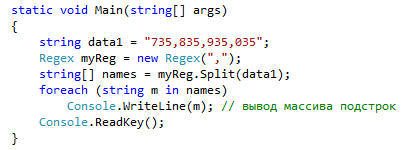


Рисунок 5 – Результат работы программы задачи 5

Таблица 2 – **Варианты заданий для задачи 5**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № варианта | Шаблон-разделитель |  | № варианта | Шаблон-разделитель |
| 1 | . |  | 21 | & |
| 2 | - |  | 22 | % |
| 3 | + |  | 23 | ^ |
| 4 | ; |  | 24 | ! |
| 5 | ! |  | 25 | + |
| 6 | @ |  | 26 | # |
| 7 | # |  | 27 | - |
| 8 | $ |  | 28 | @ |
| 9 | % |  | 29 | . |
| 10 | ^ |  | 30 | ; |
| 11 | & |  | 31 | $ |
| 12 | \* |  | 32 | & |
| 13 | ( |  | 33 | ? |
| 14 | ? |  | 34 | ( |
| 15 | / |  |  |  |
| 16 | \ |  |  |  |
| 17 | | |  |  |  |
| 18 | ~ |  |  |  |
| 19 | ) |  |  |  |
| 20 | = |  |  |  |

**3 Специальные символы**

В задачах 1 – 5 рассмотрены простые, однозначные регулярные выражения без использования спецсимволов. Существует достаточно много спецсимволов. Рассмотрим их описание в таблицах.

Таблица 3 – Классы символов

|  |  |  |  |
| --- | --- | --- | --- |
| Обозначение | Описание | Пример шаблона | Пример соответствия шаблону |
| [группа\_символов] | Любой из перечисленных в скобках символов. Используя тире можно указать диапазон символов, например, [a-f] - то же самое, что [abcdef] | [abc] | «a» в «and» |
| [^группа\_символов] | Любой символ, кроме перечисленных в скобках | [^abc] | «n», «d» в «and» |
| \d | Цифра. Эквивалентно [0-9] | \d | «2» в «data2» |
| \D | Любой символ, кроме цифр. Эквивалентно [^0-9] | \D | «y» в «2020y» |
| \w | Цифра, буква (латинский алфавит) или знак подчеркивания. Эквивалентно [0-9a-zA-Z\_] | \w | «1», «7», «с» в «1.7с» |
| \W | Любой символ, кроме цифр, букв (латинский алфавит) и знака подчеркивания. Эквивалентно [^0-9a-zA-Z\_] | \W | «.» в «1.7с» |
| \s | Пробельный символ (пробел, табуляция, перевод строки и т.п.) | \s | « » в «c sharp» |
| \S | Любой символ, кроме пробельных | \S | «c» «s» «h» «a» «r» «p» в «c sharp» |
| . | Любой символ, кроме перевода строки. Для поиска любого символа, включая перевод строки, можно использовать конструкцию [\s\S] | c.harp | «csharp» в «mycsharp» |

Таблица 4 – Символы повторения

|  |  |  |  |
| --- | --- | --- | --- |
| Обозначение | Описание | Пример шаблона | Пример соответствия шаблону |
| \* | Соответствует предыдущему элементу ноль или более раз | \d\*. | «a», «1b», «23c» в «a1b23c» |
| + | Соответствует предыдущему элементу один или более раз | \d+. | «1b», «23c » в «a1b23c» |
| ? | Соответствует предыдущему элементу ноль или один раз | \d?\D | «a», «1b», «3с» в «a1b23c» |
| {n} | Соответствует предыдущему элементу, который повторяется ровно n раз | \d{2} | «43», «54», «82» в «2,43,546,82» |
| {n,} | Соответствует предыдущему элементу, который повторяется минимум n раз | \d{2,} | «43», «546», «82» в «2,43,546,82» |
| {n,m} | Соответствует предыдущему элементу, который повторяется минимум n раз и максимум m | \d{2,} | «43», «546», «821» в «2,43,546,8212» |

Таблица 5 – Символы привязки

|  |  |  |  |
| --- | --- | --- | --- |
| Обозначение | Описание | Пример шаблона | Пример соответствия шаблону |
| ^ | Соответствие должно находиться в начале строки | ^\d{2} | «32» в «32,43,54» |
| $ | Соответствие должно находиться в конце строки или до символа \n при многострочном поиске | \d{2}$ | «54» в «32,43,54» |
| \b | Соответствие должно находиться на границе алфавитно-цифрового символа (\w) и не алфавитно-цифрового (\W) | \b\d{2} | «32», «54» в «32 a43 54» |
| \B | Соответствие не должно находиться на границе | \B\d{2} | «43» в «32 a43 54» |
| \G | Соответствие должно находиться на позиции конца предыдущего соответствия | \G\d | «3», «2», «4» в  «324.758» |

Таблица 6 – Символы выбора

|  |  |  |  |
| --- | --- | --- | --- |
| Обозначение | Описание | Пример шаблона | Пример соответствия шаблону |
| | | Работает как логическое «ИЛИ» - соответствует первому и/или второму шаблону | one|two | «one», «two» в «one two three» |
| (группа\_символов) | Группирует набор символов в единое целое для которого дальше могут использоваться + \* ? и т.д. Каждой такой группе назначается порядковый номер слева направо начиная с 1. По этому номеру можно ссылаться на группу \номер\_группы | (one)\1 | «oneone» в  «oneone onetwoone» |
| (?:группа\_символов) | Та же группировка только без назначения номера группы | (?:one){2} | «oneone» в  «oneone onetwoone» |

Таблица 7 – Служебные символы

|  |  |  |  |
| --- | --- | --- | --- |
| Обозначение | Описание | Пример шаблона | Пример соответствия шаблону |
| \t | Символ табуляции | \t |  |
| \v | Символ вертикальной табуляции | \v |  |
| \r | Символ возврата каретки | \r |  |
| \n | Символ перевода строки | \n |  |
| \f | Символ перевода страницы | \f |  |
| \ | Символ, который позволяет экранировать специальные символы, чтобы те воспринимались буквально. Например, чтобы было соответствие символу звёздочки, шаблон будет выглядеть так \\* | \d\.\d | «1.1», «1.2» в «1.1 1.2» |

**Задача 6.** Программа для проверки корректности адреса электронной почты.



Здесь перед началом строки регулярного выражения стоит символ «@», который указывает компилятору воспринимать все символы буквально. Это необходимо, чтобы корректно воспринимался символ «\».



Рисунок 6 – Результат работы программы задачи 6

**Варианты заданий для задачи 6:** проверить свои адреса электронной почты (не менее 3-х адресов).

**4 Параметры поиска**

Существует второй конструктор класса *Regex*, который принимает в качестве второго аргумента значение перечисления ***RegexOptions***. В этом перечислении есть следующие значения:

***IgnoreCase***– игнорирование регистра при поиске. Находит соответствия независимо прописными или строчными буквами в строке написано слово.

***RightToLeft***– поиск будет выполнен справа налево, а не слева направо.

***Multiline***– многострочный режим поиска. Меняет работу спецсимволов «^» и «$» так, что они соответствуют началу и концу каждой строки, а не только началу и концу целой строки.

***Singleline***– однострочный режим поиска.

***CultureInvariant***– игнорирование национальных установок строки.

***ExplicitCapture***– обеспечивается поиск только буквальных соответствий.

***Compiled***– регулярное выражение компилируется в сборку, что делает более быстрым его исполнение, но увеличивает время запуска.

***IgnorePatternWhitespace***– игнорирует в шаблоне все неэкранированные пробелы. С этим параметром шаблон «a b» будет аналогичным шаблону «ab».

***None***– использовать поиск по умолчанию.

**Задача 7.** Программа с использованием параметра поиска (игнорирование регистра).

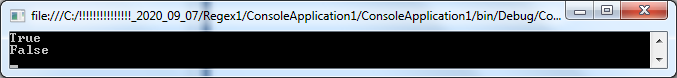
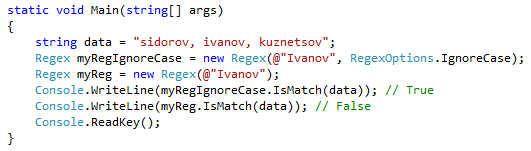


Рисунок 7 – Результат работы программы задачи 7

Если необходимо установить несколько параметров, тогда они разделяются оператором поразрядного «ИЛИ» - «|».

**Варианты заданий для задачи 7** взять из таблицы 1.

З**адача 8.** Программа для проверки правильности ввода логина. Требования к логину: корректным логином будет строка от 2-х до 10-ти символов, содержащая только латинские буквы и цифры, при этом цифра не может быть первой.

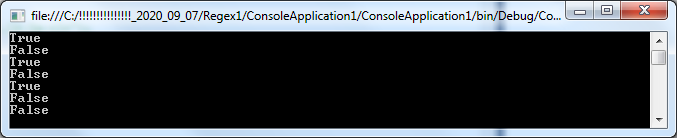
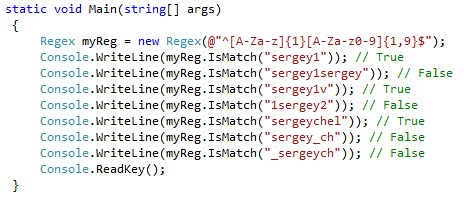


Рисунок 8 – Результат работы программы задачи 8

Таблица 8 – Варианты заданий для задачи 8

|  |  |
| --- | --- |
| № варианта | Требования к логину |
| 1 | Строка от 4 до 14 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 2 | Строка от 5 до 16 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 3 | Строка от 6 до 17 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 4 | Строка от 3 до 11 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 5 | Строка от 4 до 12 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 6 | Строка от 5 до 13 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 7 | Строка от 6 до 14 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 8 | Строка от 7 до 15 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 9 | Строка от 3 до 16 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 10 | Строка от 4 до 15 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 11 | Строка от 5 до 14 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 12 | Строка от 6 до 13 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 13 | Строка от 7 до 12 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 14 | Строка от 3 до 13 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 15 | Строка от 4 до 14 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 16 | Строка от 5 до 15 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 17 | Строка от 6 до 16 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 18 | Строка от 3 до 12 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 19 | Строка от 3 до 14 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 20 | Строка от 3 до 15 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 21 | Строка от 4 до 13 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 22 | Строка от 5 до 11 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 23 | Строка от 5 до 12 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 24 | Строка от 3 до 19 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 25 | Строка от 4 до 15 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 26 | Строка от 5 до 17 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 27 | Строка от 3 до 17 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 28 | Строка от 5 до 18 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 29 | Строка от 6 до 18 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 30 | Строка от 3 до 18 символов, содержащая только латинские и русские буквы, цифры, нижнее подчеркивание; при этом цифра и нижнее подчеркивание не могут быть первым символом |
| 31 | Строка от 4 до 16 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 32 | Строка от 6 до 19 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 33 | Строка от 7 до 12 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |
| 34 | Строка от 8 до 14 символов, содержащая только латинские и русские буквы, цифры, дефис; при этом цифра и дефис не могут быть первым символом |

**Контрольные вопросы**

1. Понятие регулярных выражений.

2. Класс *Regex*.

3. Метод *IsMatch*().

4. Метод *Match*().

5. Метод *Matches*().

6. Метод *Replace*().

7. Метод *Split*().

8. Классы символов регулярных выражений.

9. Символы повторения.

10. Символы привязки.

11. Символы выбора.

12. Служебные символы.

13. Перечисление *RegexOptions*.