

Лабораторная работа №3 «Регулярные выражения»

Для определения варианта используйте свой табельный номер, которые можно найти в ИСУ.
(Пример номера: 125598)

Задание на 60 баллов (Смайлики)

- 1) Реализуйте программный продукт на языке Python, используя регулярные выражения по варианту, представленному в таблице.
- 2) Для своей программы придумайте минимум 5 тестов. Каждый тест является отдельной сущностью, передаваемой регулярному выражению для обработки. Для каждого теста необходимо самостоятельно (без использования регулярных выражений) найти правильный ответ. После чего сравнить ответ, выданный программой, и полученный самостоятельно.
- 3) Программа должна считать количество смайликов определённого вида (вид смайлика описан в таблице вариантов) в предложенном тексте. Все смайлики имеют такую структуру:

[глаза][нос][рот].

Вариантом является различные наборы глаз, носов и ртов.

Номер в ИСУ % 6	Глаза	Номер в ИСУ % 5	Нос	Номер в ИСУ % 7	Рот
0	:	0	-	0	(
1	;	1	<	1)
2	X	2	-{	2	O
3	8	3	<{	3	
4	=			4	\
5	[5	/
				6	P

Пример смайлика: 8<{P

Необязательное задания для получения оценки «4» или «5» (позволяет набрать +18 баллов от максимального числа баллов БарС за данную лабораторную)

- 1) Реализуйте программный продукт на языке Python, используя регулярные выражения по варианту, представленному в таблице.
- 2) Для своей программы придумайте минимум 5 тестов. Каждый тест является отдельной сущностью, передаваемой регулярному выражению для обработки. Для каждого теста необходимо самостоятельно (без использования регулярных выражений) найти правильный ответ. После чего сравнить ответ, выданный программой, и полученный самостоятельно.
Пример тестов приведён в таблице.
- 3) Можно использовать циклы и условия, но основной частью решения должны быть регулярные выражения.

Номер в ИСУ % 6	Задание
0	Хайку – жанр традиционной японской лирической поэзии века, известный с XIV века.

	<p>Оригинальное японское хайку состоит из 17 слогов, составляющих один столбец иероглифов. Особыми разделительными словами – кирэдзи – текст хайку делится на части из 5, 7 и снова 5 слогов. При переводе хайку на западные языки традиционно вместо разделительного слова использую разрыв строки и, таким образом, хайку записываются как трёхстишия.</p> <p>Перед вами трёхстишия, которые претендуют на то, чтобы быть хайку. В качестве разделителя строк используются символы «/». Если разделители делят текст на строки, в которых 5/7/5 слогов, то выведите «Хайку!». Если число строк не равно 3, то выведите строку «Не хайку. Должно быть 3 строки.». Иначе выведите строку вида «Не хайку.»</p> <p>Для простоты будем считать, что слогов ровно столько же, сколько гласных, не задумываясь о тонкостях.</p> <p>Пример:</p> <table border="1" data-bbox="312 685 1484 913"> <tr> <th>Ввод</th><th>Вывод</th></tr> <tr> <td>Вечер за окном. / Еще один день прожит. / Жизнь скоротечна...</td><td>Хайку!</td></tr> <tr> <td>Просто текст</td><td>Не хайку. Должно быть 3 строки.</td></tr> <tr> <td>Как вишня расцвела! / Она с коня согнала / И князя-гордеца.</td><td>Не хайку.</td></tr> </table>	Ввод	Вывод	Вечер за окном. / Еще один день прожит. / Жизнь скоротечна...	Хайку!	Просто текст	Не хайку. Должно быть 3 строки.	Как вишня расцвела! / Она с коня согнала / И князя-гордеца.	Не хайку.
Ввод	Вывод								
Вечер за окном. / Еще один день прожит. / Жизнь скоротечна...	Хайку!								
Просто текст	Не хайку. Должно быть 3 строки.								
Как вишня расцвела! / Она с коня согнала / И князя-гордеца.	Не хайку.								
1	<p>Довольно распространённая ошибка – это повтор слова. Вот в предыдущем предложении такая допущена. Необходимо исправить каждый такой повтор.</p> <p>Повтор это – слово, один или несколько пробельных символов, и снова то же слово.</p> <p>Пример:</p> <table border="1" data-bbox="312 1171 1484 1357"> <tr> <th>Ввод</th><th>Вывод</th></tr> <tr> <td>Довольно распространённая ошибка ошибка – это лишний повтор повтор слова слова. Смешно, не не правда ли? Не нужно портить хор хоровод.</td><td>Довольно распространённая ошибка – это лишний повтор слова. Смешно, не правда ли? Не нужно портить хор хоровод.</td></tr> </table>	Ввод	Вывод	Довольно распространённая ошибка ошибка – это лишний повтор повтор слова слова. Смешно, не не правда ли? Не нужно портить хор хоровод.	Довольно распространённая ошибка – это лишний повтор слова. Смешно, не правда ли? Не нужно портить хор хоровод.				
Ввод	Вывод								
Довольно распространённая ошибка ошибка – это лишний повтор повтор слова слова. Смешно, не не правда ли? Не нужно портить хор хоровод.	Довольно распространённая ошибка – это лишний повтор слова. Смешно, не правда ли? Не нужно портить хор хоровод.								
2	<p>Дан текст. Необходимо найти в нём каждый фрагмент, где сначала идёт слово «ВТ», затем не более 4 слов, и после этого идёт слово «ИТМО».</p> <p>Для простоты будем считать словом любую последовательность букв, цифр и знаков «_» (то есть символов \w).</p> <p>Пример:</p> <table border="1" data-bbox="312 1653 1484 1767"> <tr> <th>Ввод</th><th>Вывод</th></tr> <tr> <td>А ты знал, что ВТ – лучшая кафедра в ИТМО?</td><td>ВТ лучшая кафедра в ИТМО</td></tr> </table>	Ввод	Вывод	А ты знал, что ВТ – лучшая кафедра в ИТМО?	ВТ лучшая кафедра в ИТМО				
Ввод	Вывод								
А ты знал, что ВТ – лучшая кафедра в ИТМО?	ВТ лучшая кафедра в ИТМО								
3	<p>Дан текст. Требуется найти в тексте все фамилии, отсортировав их по алфавиту.</p> <p>Фамилией для простоты будем считать слово с заглавной буквой, после которого идут инициалы.</p> <p>Пример:</p> <table border="1" data-bbox="312 2024 1484 2067"> <tr> <th>Ввод</th><th>Вывод</th></tr> </table>	Ввод	Вывод						
Ввод	Вывод								

	Студент Вася вспомнил, что на своей лекции Балакшин П.В. упоминал про старшекурсников, которые будут ему помогать: Анищенко А.А. и Машина Е.А.	Анищенко Балакшин Машина				
4	<p>Анатолий выложил пост с расписанием доп. занятий по информатике, но везде перепутал время. Поэтому нужно заменить все вхождения времени на строку (TBD).</p> <p>Время – это строка вида HH:MM:SS или HH:MM, в которой HH – число от 00 до 23, а MM и SS – число от 00 до 59.</p> <p>Пример:</p> <table><tr><td>Ввод</td><td>Вывод</td></tr><tr><td>Уважаемые студенты! В эту субботу в 15:00 планируется доп. занятие на 2 часа. То есть в 17:00:01 оно уже точно кончится.</td><td>Уважаемые студенты! В эту субботу в (TBD) планируется доп. занятие на 2 часа. То есть в (TBD) оно уже точно кончится.</td></tr></table>		Ввод	Вывод	Уважаемые студенты! В эту субботу в 15:00 планируется доп. занятие на 2 часа. То есть в 17:00:01 оно уже точно кончится.	Уважаемые студенты! В эту субботу в (TBD) планируется доп. занятие на 2 часа. То есть в (TBD) оно уже точно кончится.
Ввод	Вывод					
Уважаемые студенты! В эту субботу в 15:00 планируется доп. занятие на 2 часа. То есть в 17:00:01 оно уже точно кончится.	Уважаемые студенты! В эту субботу в (TBD) планируется доп. занятие на 2 часа. То есть в (TBD) оно уже точно кончится.					
5	<p>С помощью регулярного выражения найти в тексте все слова, в которых две гласные стоят подряд, а после этого слова идёт слово, в котором не больше 3 согласных.</p> <p>Пример:</p> <table><tr><td>Ввод</td><td>Вывод</td></tr><tr><td>Кривошеее существо гуляет по парку</td><td>гуляет</td></tr></table>		Ввод	Вывод	Кривошеее существо гуляет по парку	гуляет
Ввод	Вывод					
Кривошеее существо гуляет по парку	гуляет					

Необязательное задания для получения оценки «4» или «5» (позволяет набрать +22 балла от максимального числа баллов БаРС за данную лабораторную)

- 1) Реализуйте программный продукт на языке Python, используя регулярные выражения по варианту, представленному в таблице.
- 2) Для своей программы придумайте минимум 5 тестов.
- 3) Протестируйте свою программу на этих тестах.
- 4) Можно использовать циклы и условия, но основной частью решения должны быть регулярные выражения.

Номер в ИСУ % 5	Задание								
0	<p>Написать регулярное выражение, которое проверяет корректность email и в качестве ответа выдаёт почтовый сервер (почтовый сервер – часть email идущая после «@»).</p> <p>Для простоты будем считать, что почтовый адрес может содержать в себе буквы, цифры, «.» и «_», а почтовый сервер только буквы и «.». При этом почтовый сервер, обязательно должен содержать верхний уровень домена («.ru», «.com», etc.)</p> <p>Пример:</p> <table border="1"> <thead> <tr> <th>Ввод</th><th>Вывод</th></tr> </thead> <tbody> <tr> <td>students.spam@yandex.ru</td><td>yandex.ru</td></tr> <tr> <td>example@example</td><td>Fail!</td></tr> <tr> <td>example@example.com</td><td>example.com</td></tr> </tbody> </table>	Ввод	Вывод	students.spam@yandex.ru	yandex.ru	example@example	Fail!	example@example.com	example.com
Ввод	Вывод								
students.spam@yandex.ru	yandex.ru								
example@example	Fail!								
example@example.com	example.com								
1	<p>С помощью регулярного выражения найти в тексте слова, в которых встречается строго одна гласная буква (встречаться она может несколько раз). Пример таких слов: окно, трава, молоко, etc.</p>								

	<p>После чего данные слова требуется отсортировать по увеличению длины слова.</p> <p>Пример:</p> <table> <tr> <th>Ввод</th><th>Вывод</th></tr> <tr> <td>Классное слово – обороноспособность, которое должно идти после слов: трава и молоко.</td><td>и слов идти слово трава должно молоко обороноспособность</td></tr> </table>	Ввод	Вывод	Классное слово – обороноспособность, которое должно идти после слов: трава и молоко.	и слов идти слово трава должно молоко обороноспособность
Ввод	Вывод				
Классное слово – обороноспособность, которое должно идти после слов: трава и молоко.	и слов идти слово трава должно молоко обороноспособность				
2	<p>Студент Вася очень любит курс «Компьютерная безопасность». Однажды Васе задали домашнее задание зашифровать данные, переданные в сообщении. Недолго думая, Вася решил заменить все целые числа на функцию от этого числа. Функцию он придумал не сложную $4x^2 - 7$, где x – исходное число. Помогите Васе с его домашним заданием.</p> <p>Пример:</p> <table> <tr> <th>Ввод</th><th>Вывод</th></tr> <tr> <td>$20 + 22 = 42$</td><td>$1593 + 1929 = 7049$</td></tr> </table>	Ввод	Вывод	$20 + 22 = 42$	$1593 + 1929 = 7049$
Ввод	Вывод				
$20 + 22 = 42$	$1593 + 1929 = 7049$				
3	<p>Вывесили списки стипендиатов текущего семестра, которые представляют из себя список людей ФИО и номер группы этого человека. Вы решили подшутить над некоторыми из своих одногруппников и удалить их из списка.</p> <p>С помощью регулярного выражения найдите всех студентов своей группы, у которых инициалы начинаются на одну и ту же букву и исключите их из списка.</p> <p>Могут существовать двойные фамилии, которые тоже нужно учитывать (студенты с такими фамилиями тоже должны иметь право быть удаленными из списка стипендиатов текущего семестра).</p> <p>Пример (группа P0000):</p> <table> <tr> <th>Ввод</th><th>Вывод</th></tr> <tr> <td>Петров П.П. P0000 Анищенко А.А. P33113 Примеров Е.В. P0000 Иванов И.И. P0000</td><td>Анищенко А.А. P33113 Примеров Е.В. P0000</td></tr> </table>	Ввод	Вывод	Петров П.П. P0000 Анищенко А.А. P33113 Примеров Е.В. P0000 Иванов И.И. P0000	Анищенко А.А. P33113 Примеров Е.В. P0000
Ввод	Вывод				
Петров П.П. P0000 Анищенко А.А. P33113 Примеров Е.В. P0000 Иванов И.И. P0000	Анищенко А.А. P33113 Примеров Е.В. P0000				
4	<p>Необходимо выбрать три любых буквы и расстояние между ними. С помощью регулярного выражения нужно найти все слова (последовательность символов ограниченная пробелами), в которых встречаются эти буквы в заданной последовательности и расстояние (например, через один друг от друга).</p> <p>Пример (буквы К, Р, А через один строго в такой последовательности):</p> <table> <tr> <th>Ввод</th><th>Вывод</th></tr> <tr> <td>КоРмА КоРкА КоРчмА</td><td>КоРмА</td></tr> </table>	Ввод	Вывод	КоРмА КоРкА КоРчмА	КоРмА
Ввод	Вывод				
КоРмА КоРкА КоРчмА	КоРмА				