

Лабораторная работа №4 «Регулярные выражения»

Для определения варианта используйте свой порядковый номер по списку группы, которые можно найти в ИСУ.

Задание на 70 баллов

- 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.

Номер в ИСУ % 5	Задание								
0	<p>Хайку – жанр традиционной японской лирической поэзии века, известный с XIV века.</p> <p>Оригинальное японское хайку состоит из 17 слогов, составляющих один столбец иероглифов. Особыми разделительными словами – кирэджи – 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>Вам надо проверить, что текст является 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><tr><td>Ввод</td><td>Вывод</td></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 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><tr><td>Ввод</td><td>Вывод</td></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 любой 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="312 353 1484 472"> <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="312 725 1484 956"> <tr> <th>Ввод</th><th>Вывод</th></tr> <tr> <td>Студент Вася вспомнил, что на своей лекции Балакшин П.В. упоминал про старшекурсников, которые будут ему помогать: Анищенко А.А. и Машина Е.А.</td><td>Анищенко Балакшин Машина</td></tr> </table>	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
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 border="1" data-bbox="312 1245 1484 1435"> <tr> <th>Ввод</th><th>Вывод</th></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.				

Доп. задание №1 (+12 баллов)

- 1) Для выполнения задания скачайте текстовый документ по ссылке, соответствующей вашему варианту.

Номер в ИСУ % 30	Ссылка на текст	Описание
0 ... 9	https://drive.google.com/file/d/1j2p-EAwainLQ7dbOf2CFaXDTjQ9O7tNl/view	Гамлет
10 ... 19	https://drive.google.com/file/d/1Bo2ExJSIWc-goGx4Ton517dllr2Q8qC/view	Макбет
20 ... 29	https://drive.google.com/file/d/1YoQjz0dXSz9XFrPIIBHM_CHmJKHpYlNo/view	Ромео и Джульетта

- 2) Реализуйте программный продукт на языке Python, используя регулярные выражения, выделите все предложения из вашего текста, соответствующие критерию по вашему варианту.
- 3) Придумайте способ, как проверить корректность работы вашего регулярного выражения и проверьте его.

Номер в ИСУ % 4	Задание
0	Все предложения, в которых ровно 6 слов, одно из которых двусложное
1	Все восклицательные предложения, содержащие хотя бы одну запятую
2	Все предложения, в которых используется две и более запятых
3	Все вопросительные предложения, в которых слов меньше 6

Доп. задание №2 (+18 баллов)

- 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.

Номер в ИСУ % 6	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> <tr> <th>Ввод</th><th>Вывод</th></tr> <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> </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>трава молоко обороноспособность</td></tr> </table>	Ввод	Вывод	Классное слово – обороноспособность, которое должно идти после слов: трава и молоко.	трава молоко обороноспособность				
Ввод	Вывод								
Классное слово – обороноспособность, которое должно идти после слов: трава и молоко.	трава 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 $3x^2 + 5$, где x – исходное число. Помогите Васе с его домашним заданием.</p> <p>Пример:</p> <table border="1" data-bbox="320 353 1484 432"> <tr> <th>Ввод</th><th>Вывод</th></tr> <tr> <td>$20 + 22 = 42$</td><td>$1205 + 1457 = 5297$</td></tr> </table>	Ввод	Вывод	$20 + 22 = 42$	$1205 + 1457 = 5297$
Ввод	Вывод				
$20 + 22 = 42$	$1205 + 1457 = 5297$				
3	<p>С помощью регулярного выражения найти числа, в которых в середине числа стоит 0 и затем перенести его в конец числа.</p> <p>Пример:</p> <table border="1" data-bbox="320 616 1484 694"> <tr> <th>Ввод</th><th>Вывод</th></tr> <tr> <td>$101 + 2430123 = 2430224$</td><td>$110 + 2431230 = 2432240$</td></tr> </table>	Ввод	Вывод	$101 + 2430123 = 2430224$	$110 + 2431230 = 2432240$
Ввод	Вывод				
$101 + 2430123 = 2430224$	$110 + 2431230 = 2432240$				
4	<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 border="1" data-bbox="320 878 1484 956"> <tr> <th>Ввод</th><th>Вывод</th></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				
5	<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 одноклассников 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>Пример (группа P000):</p> <table border="1" data-bbox="320 1288 1484 1478"> <tr> <th>Ввод</th><th>Вывод</th></tr> <tr> <td>Петров П.П. P000 Анищенко А.А. P33113 Примеров Е.В. P000 Иванов И.И. P000</td><td>Анищенко А.А. P33113 Примеров Е.В. P000</td></tr> </table>	Ввод	Вывод	Петров П.П. P000 Анищенко А.А. P33113 Примеров Е.В. P000 Иванов И.И. P000	Анищенко А.А. P33113 Примеров Е.В. P000
Ввод	Вывод				
Петров П.П. P000 Анищенко А.А. P33113 Примеров Е.В. P000 Иванов И.И. P000	Анищенко А.А. P33113 Примеров Е.В. P000				