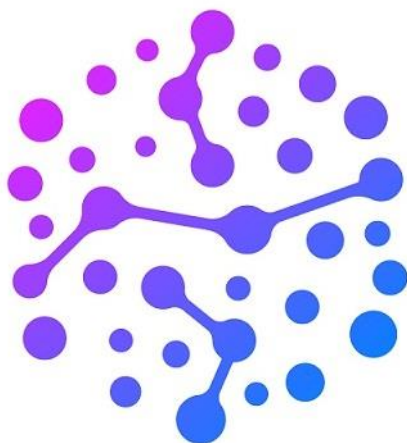


ЗАДАЧИ

ВСЕРОССИЙСКОЙ ОЛИМПИАДЫ ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ – 2021 –



УСЛОВИЯ ЗАДАЧ ОТБОРОЧНОГО ЭТАПА ВСЕРОССИЙСКОЙ ОЛИМПИАДЫ ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ – 2021

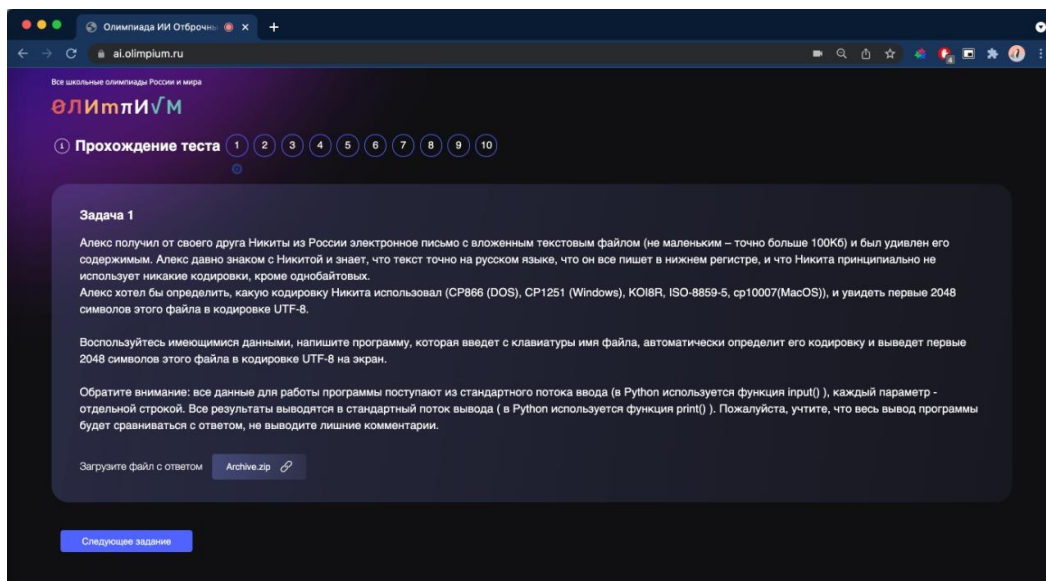
Задача № 1.1

Алекс получил от своего друга Никиты из России электронное письмо с вложенным текстовым файлом (не маленьким – точно больше 100 Кб) и был удивлен его содержимым. Алекс давно знаком с Никитой и знает, что текст точно на русском языке, что Никита все пишет в нижнем регистре и принципиально не использует никакие кодировки, кроме однобайтовых.

Алекс хотел бы определить, какую кодировку использовал Никита (CP866 (DOS), CP1251 (Windows), KOI8R, ISO-8859-5, cp10007 (MacOS)), и увидеть первые 2048 символов этого текста в кодировке UTF-8.

Воспользуйтесь имеющимися данными и напишите программу, которая будет автоматически определять кодировку и выводить первые 2048 символов текста в кодировке UTF-8.

Обратите внимание: все данные для работы программы поступают из стандартного потока ввода (в Python используется функция `input()`), каждый параметр – отдельной строкой. Все результаты выводятся в стандартный поток вывода (в Python используется функция `print()`). Пожалуйста, учтите, что весь вывод программы будет сравниваться с ответом, не выводите лишние комментарии.



Задача № 1.2

Во фрагменте текста (текстовый файл) для красоты решили записать все натуральные числа (годы, номера королей) римскими числами.

Известно, что в тексте встречаются числа от 1 до 3100.

Напишите программу, которая введет с клавиатуры имя текстового файла в кодировке UTF-8, переведет все такие числа в римскую запись (по правилам Шварцмана) и выведет текст с заменой их записи на экран.

Учтите, что даты и дробные числа преобразовывать нельзя, хотя цифры там есть.

Напоминаем:

- Римские целые числа записываются буквами I, V, X и некоторыми другими;
- Числа, написанные с уменьшением значения слева направо, суммируются (VI – 6), а с уменьшением значения справа налево вычитаются (IX – 9);
- количество символов в записи должно быть минимальным;
- больше 3 одинаковых символов подряд писать нельзя.

Обратите внимание: все данные для работы программы поступают из стандартного потока ввода (в Python используется функция `input()`), каждый параметр – отдельной строкой. Все результаты выводятся в стандартный поток вывода (в Python используется функция `print()`). Пожалуйста, учтите, что весь вывод программы будет сравниваться с ответом, не выводите лишние комментарии.

Задача № 1.3

На уроке литературы возник спор: какие слова в текстах на русском языке используются чаще всего. Для разрешения спора решили провести эксперимент – посчитать слова в большом литературном произведении. С учетом объема решили написать программу для такой обработки текста. Произведение для анализа дают в виде текстового файла на

русском языке в кодировке UTF-8, записанного по правилам набора текстов (возможно, с переносами).

Напишите программу, которая введет с клавиатуры имя файла, посчитает частоту слов в этом файле и выведет на экран через пробел 50 самых частых слов по убыванию частоты.

Обратите внимание: все данные для работы программы поступают из стандартного потока ввода (в Python используется функция `input()`), каждый параметр – отдельной строкой. Все результаты выводятся в стандартный поток вывода (в Python используется функция `print()`). Пожалуйста, учтите, что весь вывод программы будет сравниваться с ответом, не выводите лишние комментарии.

Задача № 1.4

Есть текстовый файл, в котором встречаются натуральные шестнадцатеричные числа. Их запись начинается с `0x` и использует большие латинские буквы (например, `0x1A`). Других чисел в файле нет. К сожалению, некоторые числа записаны неправильно – пропущено `0x`. Иногда после `0x` записано не число. Напишите программу, которая введет с клавиатуры имя текстового файла и выведет на экран из текста шестнадцатеричные числа, добавив, где нужно, `0x`.

Обратите внимание: все данные для работы программы поступают из стандартного потока ввода (в Python используется функция `input()`), каждый параметр – отдельной строкой. Все результаты выводятся в стандартный поток вывода (в Python используется функция `print()`). Пожалуйста, учтите, что весь вывод программы будет сравниваться с ответом, не выводите лишние комментарии.

Задача № 1.5

Работая редактором, Вы получаете от авторов тексты на русском языке, набранные не очень аккуратно. Текст состоит из абзацев, но используются переносы слов, абзац выделяется разным количеством символов, неаккуратно оформлены знаки препинания... В производство такой файл отдавать нельзя, исправлять вручную тяжело, особенно если текст большой.

Напишите программу, которая введет с клавиатуры имя текстового файла, приведет текст в соответствие с правилами машинописного набора и выведет результат на экран.

Правила набора:

- абзац начинается с 3 пробелов;
- все строки абзаца не имеют переносов;
- перед знаком препинания нет пробелов, но обязательно есть один пробел после;
- кавычки и скобки замыкают фрагменты текста (например, так);
- повторы знаков препинания (за исключением точек, вопросительного и восклицательного знака, не более 3 раз) не допускаются.

Обратите внимание: все данные для работы программы поступают из стандартного потока ввода (в Python используется функция `input()`), каждый параметр – отдельной строкой. Все результаты выводятся в стандартный поток вывода (в Python используется функция `print()`). Пожалуйста, учтите, что весь вывод программы будет сравниваться с ответом, не выводите лишние комментарии.

Задача № 1.6

С клавиатуры вводят два имени текстовых файлов ($>1M <10M$). Файлы содержат целые числа.

Написать программу, которая выведет через пробел числа, которые есть во втором файле, но которых нет в первом, не более чем за 5 секунд.

Обратите внимание: все данные для работы программы поступают из стандартного потока ввода (в Python используется функция `input()`), каждый параметр – отдельной строкой. Все результаты выводятся в стандартный поток вывода (в Python используется функция `print()`). Пожалуйста, учтите, что весь вывод программы будет сравниваться с ответом, не выводите лишние комментарии.

Задача № 1.7

Лена и Вася собираются в Красивый Город на шестнадцать дней и имеют на руках некоторую сумму денег (пока они не знают, сколько). Ребята хотят посетить самые интересные места. О каждом месте они знают, насколько оно интересно (в баллах) и сколько стоит его посетить (в деньгах); список есть в файле, имя которого вводится с клавиатуры. К сожалению, они могут посещать только два места в день.

Разработайте программу, которая получит с клавиатуры максимальную сумму, имя файла-справочника и выведет список мест, которые им стоит посетить: максимально интересные (в сумме баллов) и допустимые по деньгам. Места в списке перечислить через точку с запятой, в том числе в конце списка.

Обратите внимание: все данные для работы программы поступают из стандартного потока ввода (в Python используется функция `input()`), каждый параметр – отдельной строкой. Все результаты выводятся в стандартный поток вывода (в Python используется функция `print()`). Пожалуйста, учтите, что весь вывод программы будет сравниваться с ответом, не выводите лишние комментарии.

Задача № 1.8

У телефонной компании есть большое количество 10-значных телефонных номеров; заранее неизвестно, сколько их, но почти все возможные номера используются. Память телефонной станции ограничена, поэтому на хранение номеров в «рабочем» режиме отводится не более 1,4 Мб.

Написать программу, которая прочитает и обработает файл с телефонными номерами (текстовый, один номер – одна строка, ведущие нули не записываются) и выдаст список свободных (то есть отсутствующих в файле) номеров не более чем за 15 секунд.

Обратите внимание: все данные для работы программы поступают из стандартного потока ввода (в Python используется функция `input()`), каждый параметр – отдельной строкой. Все результаты выводятся в стандартный поток вывода (в Python используется функция `print()`).

Пожалуйста, учтите, что весь вывод программы будет сравниваться с ответом, не выводите лишние комментарии.

Задача № 1.9

Два автора прислали в издательство свои произведения, написанные в жанре фантастики и включающие множество уникальных слов. Требуется выявить совпадения по уникальным словам. Программисту издательства поставили эту задачу в следующей формулировке. Даны имена двух больших (до 10 Мбайт) текстовых файлов в кодировке UTF-8 с текстом на русском языке.

Написать программу, которая выдаст списки 50 уникальных слов первого файла (по убыванию частоты) и 50 уникальных слов второго файла (по убыванию частоты), используя не более чем 5 Мбайт памяти. Слова вывести через один пробел.

Обратите внимание: все данные для работы программы поступают из стандартного потока ввода (в Python используется функция `input()`), каждый параметр – отдельной строкой. Все результаты выводятся в стандартный поток вывода (в Python используется функция `print()`). Пожалуйста, учтите, что весь вывод программы будет сравниваться с ответом, не выводите лишние комментарии.

Задача № 1.10

Для учебно-исследовательского проекта Вася пишет очень длинное математическое выражение, в котором много разных скобок. Скобки – это один символ на открытие и один символ на закрытие (например: / и / или [и]). Вася иногда ошибается в таких выражениях – неправильно ставит скобки, причем может допустить несколько ошибок в одной формуле.

Написать программу, которая получает строку из открывающих частей скобок, строку из закрывающих частей скобок и само выражение.

Программа должна проверить выражение и выдать такие оценки:

- Все в порядке

- Вы не все закрывающие скобки ввели
- Вы не все открывающие скобки ввели
- Неравное количество открытых и закрытых скобок
- Нельзя закрывать открывающей скобкой
- В выражении перепутан порядок скобок ... и ...
(например: В выражении перепутан порядок скобок { и])
- Не все скобки вида ... закрыты
- Не все скобки вида ... открыты

УСЛОВИЯ ЗАДАЧ ОСНОВНОГО ЭТАПА ВСЕРОССИЙСКОЙ ОЛИМПИАДЫ ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ – 2021

Задача № 2.1

Дана фотография одинокого яблока крупно на очень светлом или очень темном фоне в формате JPEG (образец прилагается). Имя файла с фотографией вводится с клавиатуры.

Какое это яблоко – зеленое, желтое, белое (очень светло-желтое или светло-зеленое) или красное?

Ответ вывести названием цвета (одно слово) на русском языке в нижнем регистре.

Задача № 2.2

Есть фото участка стенки до начала покраски и во время покраски, в формате JPEG.

Написать программу, которая получит с клавиатуры название файла с участком непокрашенной стены, название файла с покрашенной стеной и определит, сколько процентов стены уже покрашено (то есть где появилась краска). Учтите, что ровным и точным цвет не бывает.

Ответ вывести в целых процентах, возможно отклонение до 2% включительно.

Задача № 2.3

Дмитрий Дремов (<https://dremovd.github.io/receipt-data/>) создал набор данных из кассовых чеков.

Мы немного изменили этот набор, не меняя его структуру и статистику покупок.

Напишите программу, которая проанализирует набор товаров в чеках (файл CSV в кодировке windows-1251 с разделителем точка с запятой, имя файла вводится с клавиатуры)

и предскажет по введенному товару – какой товар (или товары) скорее всего купят вместе с ним?

Товары вывести через запятую.

При анализе и вводе названия продуктов «очищаются» – переписываются в нижнем регистре и из них исключены все цифры и знаки препинания, все слова в названии написаны через 1 пробел.

Например: мозырское молоко паст л ф п мозырские мл

Задача № 2.4

Имеется база отзывов на фильмы – в виде общего каталога с текстовыми файлами, разделенным внутри на три каталога – ros, neg и neu.

В каждом из этих каталогов несколько (>10) текстовых файлов на русском языке в кодировке UTF-8.

Выведите через 1 пробел список из 50 самых информативных слов – то есть тех, которые лучше всего позволяют различать, к какой категории относятся отзывы.

Информативность слов оценивают с помощью критерия взаимной информации

(https://ru.abcdef.wiki/wiki/Mutual_information#Conditional_mutual_information)

Задача № 2.5

Aditya Kadiwal собрал набор данных о нормальных и мошеннических транзакциях с банковскими картами:
<https://www.kaggle.com/adityakadiwal/credit-card-fraudulent-transactions>

Напишите программу, которая введет с клавиатуры имя файла с такими данными, а после этого по аналогичным данным сможет принять решение – разрешать транзакцию или отправить на проверку человеком?..

В случае разрешения вывести на экран слово LEGIT, при отправке на проверку FRAUD

Данные для проверки вводятся с клавиатуры в виде строки, с разделителем точкой с запятой, также как в CSV-файле, за исключением последнего поля (TRN_TYPE)

УСЛОВИЯ ЗАДАЧ ЗАКЛЮЧИТЕЛЬНОГО ЭТАПА ВСЕРОССИЙСКОЙ ОЛИМПИАДЫ ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ – 2021

Задача № 3.1

Мария собрала из разных источников много отзывов на фильмы. Каждый отзыв хранится в отдельном текстовом файле, в кодировке UTF-8. Кроме отзыва файл ничего не содержит (оценок в баллах там тоже нет).

Мария устала сортировать отзывы вручную, и желает получить программу, которая будет:

1. Вводить с клавиатуры имя текстового файла
2. Выводить в стандартный вывод тип отзыва: – положительный, отрицательный или нейтральный. Результат надо вывести на экран одним словом – “positive”, “negative”, “neutral”

Мария отсортировала 200 отзывов и на них проверит работу программы. Результат оценки 200 файлов будет сравниваться с эталоном.

Итоговая оценка программы вычисляется на основе метрики F1. Программа получит $F1 * 100$ баллов (то есть если метрика $F1 = 0.7356687$, в рейтинг пойдет оценка 73.56687)

Для обучения модели можно использовать опубликованный Михаилом Клеминым набор данных (отзывы с сайта «Кинопоиск»): <https://www.kaggle.com/mikhailklemin/kinopoisks-movies-reviews>

Задача № 3.2

Два приятеля поспорили, кому из них во время путешествий удалось сделать лучшую фотографию звездного неба. Лучшим считается тот снимок, на котором запечатлено больше звезд.

Взглянув на фотографии, ребята осознали, что вручную посчитать количество звезд на снимках будет сложно и им нужна программа для такого подсчета.

Программа должна:

- ввести имя файла в формате JPEG;
- вывести количество звезд на снимке.

По условиям спора необходимо сравнить фотографии, которые были сделаны ребятами с поверхности Земли и не подвергались художественной обработке. Сами снимки вам не показывают, но они примерно такие:

<https://www.schaman.ru/foto/2016/010/03.jpg>

<http://www.fotoplex.ru/photos/andrey/night/i-372300.jpg>

https://studiya-f.ru/wp-content/uploads/2019/05/Kak_fotografirovat_zvezdnoe_nebo_v_azhnye_sovety_i_rekomendacii_1.jpg

Прятели понимают, что точный подсчет звезд на фотографиях невозможен. Договорились, что будут считать только различимые отдельные звезды (то есть видимую глазом на фотографии точку, млечный путь, там, где уже нельзя выделить звезды, на звезды делить не надо). Собрали 10 фотографий и посчитали на них звезды вручную (возможно, ошиблись – но не намного). Разрешение фотографий – от 599*533 до 2560*1600.

На этих фотографиях будут проверять программу: ошибка в 50% и более недопустима (это не может рассматриваться как результат). Поэтому за отклонение количества в 50% и более программа получает по тесту 0 баллов, при отсутствии отклонения – 10 баллов за каждую проверенную фотографию.

Программа получает из стандартного ввода имя файла с фотографией и выдает на стандартный вывод целое число – количество звезд. Программа принимается для проверки, если за подсчет звезд на тестовой фотографии получает хотя бы один балл.

ВСЕРОССИЙСКАЯ ОЛИМПИАДА ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ

ЗАДАЧИ финала олимпиады

№1.
ЗАДАЧА НА ОЦЕНКУ ТОНАЛЬНОСТИ ВЫСКАЗЫВАНИЯ
 («обработка текстов»)

«Понятно что... супергер... на...
 Отличный фильм!
 ...вообще...
 ...рсно. Да и в...
 целом мешанина...
 какая-то...»

«...ема с... фильма...
 ...вая и...
 ...исаживается...
 ...из...
 ...не...
 ...ора...
 ...а...»

«...чис...
 ...диало...
 ...за их...
 ...очень...
 ...арс...
 ...см...
 ...сучно...»

«...Во время пр...
 ...чувствова...
 ...себя...
 ...Бенском, было...
 ...полное погружение...
 ...это нечто!»

«Два приятеля поспорили, кому из них во время путешествий удалось сделать лучшую фотографию звездного неба. Лучшим считается тот снимок, на котором запечатлено больше звезд...»

№2.
ЗАДАЧА НА СЕГМЕНТИРОВАНИЕ ИЗОБРАЖЕНИЙ
 («машинное зрение»)

СОСТАВ МЕТОДИЧЕСКОЙ КОМИССИИ ВСЕРОССИЙСКОЙ ОЛИМПИАДЫ ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ

- 2021 -

1. **Григорьев Сергей Георгиевич (председатель)** – доктор технических наук, профессор, член-корреспондент Российской академии образования:
 - профессор департамента информатики, управления и технологий ГАОУ ВО МГПУ;
 - Почетный работник высшего профессионального образования Российской Федерации;
 - Почетный доктор КазНПУ им. Абая (г. Алма-Ата, Казахстан);
 - награжден Почетной грамотой Министерства образования и науки Российской Федерации, наградами Российской академии образования, Департамента образования города Москвы, Министерства образования и науки Республики Марий-Эл, МГПУ;
 - член Научно-методического совета по информатике и ИКТ ФГБНУ «ФИПИ»;
 - председатель программных комитетов Международной конференции «Инфостратегия», международного конгресса конференций «Информационные технологии в образовании» (ИТО);
 - автор свыше 350 публикаций, в т.ч. учебников и учебно-методических пособий для образовательных организаций. Главный редактор журнала «Информатика и образование», журнала «Вестник МГПУ. Серия «Информатика и информатизация образования»», член редакционных коллегий журналов «Вестник РУДН. Серия «Информатизация образования»», «Научно-технические библиотеки».

2. **Самылкина Надежда Николаевна** – кандидат педагогических наук, доцент:
 - профессор кафедры теории и методики обучения математики и информатики Института математики и информатики ФГБОУ ВО «Московский педагогический государственный университет»,
 - эксперт олимпиад РСОШ Минобрнауки России, член оргкомитета Международной олимпиады по креативному программированию (2017-2020 г.)
 - автор УМК по информатике углубленного уровня.

3. **Калинин Илья Александрович** – кандидат педагогических наук, доцент:
 - начальник управления информатизации, доцент кафедры международной информационной безопасности Института информационных наук ФГБОУ ВО МГЛУ,
 - автор УМК по информатике углубленного уровня.