

# Задания для работы с коллекциями и контейнерами

Для решения задачи необходимо разработать подходящий класс или классы. В классе должны быть только те поля, которые необходимы для решения задачи, лишняя информация не должна храниться в объектах, даже если она присутствует в файлах с исходными данными. Все элементы классов должны иметь минимально возможную область доступа. Каждый метод класса должен выполнять только одну операцию. Во всех случаях, когда возможно использование библиотечных классов Java API, они должны быть использованы.

Для решения задачи необходимо использовать подходящие коллекции или контейнеры, обязательно использовать лямбда-выражения и Stream API.

Все ошибочные ситуации должны корректно обрабатываться программой! Если во входном файле встречается ошибка и без этих данных продолжение работы невозможно, то необходимо завершить работу программы с соответствующим сообщением в консоли.

В ряде задач требуется использовать для вывода формат json. Прочитать о нём и библиотеках Java, которые помогают анализировать данные в этом формате можно в статье [Руководство Java JSON Processing API \(JSONP\)](#).

## Оценка 3

Для получения оценки 3 задача должна быть решена полностью.

**Задача 1.** В файле `data_anime.txt` содержится информация о фильмах и сериалах в стиле аниме: уникальный идентификатор, полное название аниме, жанр (список жанров этого аниме, разделенных запятыми), тип, количество серий (1, если фильм), рейтинг - средняя оценка этого аниме от 0 до 10, количество участников сообщества в этом аниме. Каждый из семи параметров расположен в отдельной строке, в следующих семи строках - информация о следующем объекте.

Выведите на консоль список всех сериалов жанров `Fantasy` и `Samurai`, отсортированных сначала по жанру, затем по названию. Для списка найдите среднее количество участников сообщества, три самых коротких и три самых длинных по количеству серий аниме. Выведите в текстовый файл информацию об объектах списка в виде таблицы со столбцами: название, жанр, количество серий. Данные в таблице должны быть упорядочены по названию.

**Задача 2.** В файле `data_auto.txt` содержится информация о продаваемых автомобилях: индекс, цена автомобиля в объявлении, марка автомобиля, модель автомобиля, год выпуска автомобиля, техническое состояние автомобиля (`clean vehicle` — хорошее, `salvage insurance` — отремонтированный после серьёзной поломки), количество миль, пройденных автомобилем, цвет автомобиля, идентификационный номер автомобиля (набор из 17 символов — цифр и заглавных букв), номер партии выпущенных автомобилей. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках - информация о следующем объекте.

Выведите на консоль и в текстовый файл список автомобилей по заданному пользователем году выпуска, отсортированных сначала по марке автомобиля, затем по пробегу. Для этого списка выведите среднюю цену автомобилей каждого технического состояния отдельно. Определите самый часто встречающийся цвет автомобилей этого списка и самый редкий.

**Задача 3.** В файле `data_book.txt` содержится информация о книгах: индекс, название, автор книги (несколько авторов разделяются символом `,`), средний рейтинг книги от 0 до 5, уникальный

идентификационный номер, международный стандартный номер книги ISBN, код языка, количество страниц, общее количество оценок, полученных книгой, общее количество рецензий на текст книги. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках - информация о следующем объекте.

Определите количество и долю книг, написанных в соавторстве, выведите на консоль список книг с тремя соавторами, отсортированный по количеству страниц. Для автора, заданного пользователем, определите количество и средний рейтинг его книг, присутствующих в файле, в отдельный файл выведите три самых рейтинговых книги этого автора.

**Задача 4.** В файле `data_course.txt` содержится информация об онлайн-курсах: идентификатор курса, название курса, URL курса, курс бесплатный (False) или платный (True), стоимость курса (0, если курс бесплатный), текущее количество слушателей, количество отзывов, количество лекций, сложность курса, продолжительность курса в месяцах. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль список всех бесплатных курсов, отсортированный по возрастанию длительности. Выведите в текстовый файл информацию о платных курсах в заданном пользователем диапазоне стоимости в виде таблицы со столбцами: название, URL курса, количество лекций, стоимость. Данные в таблице должны быть упорядочены по названию.

**Задача 5.** В файле `data_game.txt` содержится информация о компьютерных играх: идентификатор, название игры, платформа, год выпуска, жанр, организация-разработчик, цена в Северной Америке, цена в Европе, цена в Японии, цена в других странах. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл информацию о названии и разработчике самой дешёвой и самой дорогой игре в Северной Америке и в Европе, среднюю цену игр в Северной Америке, в Европе, в Японии и в других странах в году, заданном пользователем, список разработчиков 20 самых дорогих игр в заданном году.

**Задача 6.** В файле `data_journals.txt` содержится информация о научных журналах: название, импакт-фактор (международный рейтинг), категория, количество цитирований статей журнала за последние три года, количество опубликованных статей за последние три года, процентное соотношение цитирования по сравнению с предыдущим годом, коэффициент цитирования на основе общего количества цитирований в предметной области, коэффициент цитирования в рейтинговых журналах, название издателя журнала. Каждый из девяти параметров расположен в отдельной строке, в следующих девяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл список журналов, отсортированный сначала по издательству, а затем по импакт-фактору. Определите общее количество статей, напечатанное в издательстве, выбранном пользователем из предложенного списка.

**Задача 7.** В файле `data_weather.txt` содержится информация о погоде каждого дня 2009 – 2016 гг.: дата наблюдения, название метеостанции, минимальная температура в градусах Цельсия, максимальная температура в градусах Цельсия, количество осадков, зафиксированных за день в мм., так называемое испарение на сковороде за 24 часа, количество часов яркого солнечного света в день, направление сильнейшего порыва ветра, скорость (км/ч) самого сильного порыва ветра, направление ветра в 9 утра (NA — значение параметра не определено). Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл для каждого года количество дней с дождём, количество дней с сильным дождём (сильным — называется дождь, если его выпадает от 15 до 49 мм), количество дней, когда среднесуточная температура попадала в диапазон, заданный пользователем, дату с самым сильным ветром, даты трёх самых жарких и трёх самых холодных дней.

**Задача 8.** В файле `data_weather_australia.txt` содержится информация о погоде первых восьми месяцев 2009 года в Австралии, наблюдаемой на двух метеостанциях: дата наблюдения, название метеостанции, минимальная температура в градусах Цельсия, максимальная температура в градусах Цельсия, количество осадков, зафиксированных за день в мм., так называемое испарение на сковороде за 24 часа, количество часов яркого солнечного света в день, направление сильнейшего порыва ветра, скорость (км/ч) самого сильного порыва ветра, направление ветра в 9 утра (NA — значение параметра не определено). Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл информацию об осадках в те дни, когда разница между среднесуточной температурой на этих станциях превышала значение, заданное пользователем.

Определите название метеостанции, где наблюдалось наибольшее суммарное количество осадков и выведите дату, когда количество осадков было наибольшим, для сравнения выведите количество осадков в этот же день на другой метеостанции.

**Задача 9.** В файле `data_chess_player.txt` содержится информация о лучших шахматных игроках: дата составления рейтинга, рейтинг (занимаемое место), имя (фамилия и имя разделённые запятой и пробелом), звание (`g` — Grand Master, `wg` — Woman Grand Master), страна, в которой родился, стандартный шахматный рейтинг (выше 2400 — Grand Master), количество сыгранных игр за некоторый период, год рождения. Каждый из восьми параметров расположен в отдельной строке, в следующих восьми строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл списки игроков страны, заданной пользователем, отсортированные по стандартному рейтингу и по возрасту. Определите самого старшего игрока и страну его рождения и выведите на консоль список из трёх лучших по рейтингу игроков этой страны и общее количество игр, сыгранных игроками этой страны.

**Задача 10.** В файле `data_countries_world.txt` содержится информация о странах: название, описание региона, численность населения, площадь в квадратных милях (1 миля примерно 1,6 км.), плотность населения на одну квадратную милю, отношение длины побережья к площади (береговая линия), миграция в год среднее количество на 1000 человек населения, младенческая смертность на 1000 рождений, ВВП на душу населения, грамотность в процентах от всего населения. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл список всех стран, у которых есть выход к морю, отсортированный по убыванию площади. Для самой густонаселённой страны из этого списка определите регион и выведите на консоль количество стран этого региона без выхода к морю. По выбору пользователя найдите в отобранном списке страны с максимальными и минимальным ВВП или страны с максимальной и минимальной грамотностью.

**Задача 11.** В файле `data_diamond.txt` содержится информация о бриллиантах (часть параметров связаны с физическими размерами: длиной  $x$ , шириной  $y$ , высотой  $z$ ): индекс, вес в каратах, качество огранки (в порядке увеличения Fair, Good, Very Good, Premium, Ideal), цвет (буква латинского алфавита от лучшего D до худшего J), прозрачность (в порядке от лучшего к худшему: FL, IF, VVS1, VVS2, VS1, VS2, SI1, SI2, I1, I2, I3), глубина (высота алмаза делённая на его средний диаметр в процентах  $\frac{z}{(x+y)/2} \times 100 = \frac{200z}{x+y}$ ), ширина алмаза в процентах от его среднего диаметра, цена бриллианта, длина  $x$  в мм, ширина  $y$  в мм. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл списки из десяти самых дорогих и самых дешёвых бриллиантов. Из этих списков пользователь выбирает объект, после чего на консоль выводится список бриллиантов того же цвета, но большего размера как по длине, так и по ширине, список упорядочивается по возрастанию веса.

**Задача 12.** В файле `data_meteorite.txt` содержится информация о метеоритах: название, идентификатор, тип (`valid` — обычный, `relict` — метеорит, сильно разрушенный погодными условиями), класс (классификация зависит от физических, химических и некоторых других свойств метеорита), масса в граммах, наблюдалось ли падение (`Fell` — наблюдалось падение метеорита, `Found` — падения метеорита не наблюдалось), год обнаружения метеорита, широта падения метеорита, долгота падения метеорита, геолокация в виде точки с парой координат: широтой и долготой. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Для заданного пользователем диапазона лет выведите на консоль и в текстовый файл долю разрушенных метеоритов, долю метеоритов, у которых наблюдалось падение, среднюю массу, список десяти самых лёгких метеоритов, упорядоченный по названию.

## Оценка 3 — 4

Для получения оценки 3 нужно решить всё, кроме дополнительного задания. Для получения оценки 4 задача должна быть решена полностью.

**Задача 13.** В файле `data_anime.txt` содержится информация о фильмах и сериалах в стиле аниме: уникальный идентификатор, полное название аниме, жанр (список жанров этого аниме,

разделенных запятыми), тип, количество серий (1, если фильм), рейтинг - средняя оценка этого аниме от 0 до 10, количество участников сообщества в этом аниме. Каждый из семи параметров расположен в отдельной строке, в следующих семи строках - информация о следующем объекте.

Для жанра, заданного пользователем, выведите на консоль и в текстовый файл отдельно список всех фильмов и отдельно список сериалов, отсортированные по названию. Для сериалов найдите 20 лучших по рейтингу, количество участников сообщества в которых не более 200 000. Для фильмов выведите 10% лучших по рейтингу относительно общего количества фильмов, отсортированных по количеству участников сообщества.

Дополнительно составьте списки аниме каждого жанра, упорядоченные сначала по типу, затем по рейтингу в порядке возрастания. При выводе каждого списка сначала сообщите описание соответствующего жанра, количество аниме в списке, диапазон количества участников сообществ.

**Задача 14.** В файле `data_book.txt` содержится информация о книгах: индекс, название, автор книги (несколько авторов разделяются символом /), средний рейтинг книги от 0 до 5, уникальный идентификационный номер, международный стандартный номер книги ISBN, код языка, количество страниц, общее количество оценок, полученных книгой, общее количество рецензий на текст книги. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках - информация о следующем объекте.

Выведите на консоль и в текстовый файл список американских книг (код языка `en-US`), отсортированный по суммарному количеству оценок и рецензий. Из этого списка выведите 10 самых тонких и 10 самых толстых книг. Найдите книгу по названию, введённому пользователем, и выведите список остальных книг автора этой книги, если авторов несколько, выведите список книг первого автора.

Дополнительно создайте списки книг с двумя соавторами, тремя соавторами и т.д. Для каждой книги списка определите соавтора с наибольшим средним количеством оценок на одну его книгу, если в исходных данных у всех соавторов нет других книг, укажите эту информацию, результаты запишите в отдельные текстовые файлы для каждого списка.

**Задача 15.** В файле `data_chocolate.txt` содержится информация о сортах тёмного шоколада и их экспертной оценке: страна-экспортёр какао-бобов, название фирмы экспортёра, название марки шоколада, индекс экспертизы, год экспертизы, процентное содержание какао, страна, в которой выпускается сам шоколад указанной марки, экспертная оценка. Каждый из восьми параметров расположен в отдельной строке, в следующих восьми строках - информация о следующем объекте.

Выведите на консоль информацию о лучшей марке шоколада с самым высоким содержанием какао и с самым низким: название, содержание какао, страну-экспортёра какао-бобов, страну-производителя шоколада. В текстовый файл выведите количество и список марок шоколада с заданным содержанием какао, которые в заданном пользователем году получили экспертную оценку не ниже 3.

Дополнительно в текстовый файл выведите список стран-производителей шоколада, отсортированных по возрастанию средней экспертной оценки выпускаемого шоколада и для каждой страны определите лучшую страну экспортёра какао-бобов.

**Задача 16.** В файле `data_course.txt` содержится информация об онлайн-курсах: идентификатор курса, название курса, URL курса, курс бесплатный (`False`) или платный (`True`), стоимость курса (0, если курс бесплатный), текущее количество слушателей, количество отзывов, количество лекций, сложность курса, продолжительность курса в месяцах. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль списки всех курсов, длительностью не более трёх месяцев, отдельно платные и отдельно бесплатные. Среди платных найдите 10 самых популярных по количеству слушателей курсов и выведите информацию о них в текстовый файл, отсортированную по цене. Для каждого из бесплатных курсов определите среднее количество лекций в неделю и выведите на консоль список, упорядоченный по возрастанию этого среднего.

Дополнительно для каждого уровня сложности выведите списки платных и бесплатных курсов, упорядоченные по длительности, количеству лекций или количеству слушателей, в зависимости от желания пользователя.

**Задача 17.** В файле `data_game.txt` содержится информация о компьютерных играх: идентификатор, название игры, платформа, год выпуска, жанр, организация-разработчик, цена в Северной Америке, цена в Европе, цена в Японии, цена в других странах. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Определите разработчика самой дорогой игры и подсчитайте общее количество игр этого разработчика и их среднюю стоимость в регионе, выбранном пользователем, выведите информацию на консоль. В текстовый файл выведите информацию об играх заданной пользователем платформы: название, год выпуска, цена в Европе, цена в Японии, отсортированную по году выпуска.

Дополнительно выведите на консоль и в текстовый файл для каждой платформы количество игр, среднюю стоимость, названия трёх самых дорогих в Европе и в Японии и такую же информацию в отдельных файлах для каждого года из диапазона, заданного пользователем.

**Задача 18.** В файле `data_olimp.txt` содержится информация об олимпийских спортсменах: индекс, имя, пол (М — мужской или F — женский), возраст — целое число, высота в сантиметрах, вес в килограммах (если данные не известны — NA), название команды, трехбуквенный код Национального олимпийского комитета, год и сезон олимпийских игр, год. Каждый из девяти параметров расположен в отдельной строке, в следующих девяти строках — информация о следующем объекте.

Определите самого старого олимпийского игрока и выведите на консоль и в текстовый файл название и список членов его команды в соответствующем году. Для мужчин и женщин определите и выведите на консоль минимальный, максимальный, средний рост, вес и возраст. Для каждого сезона определите и выведите на консоль общее количество участвовавших в нём мужчин и женщин.

Дополнительно выведите на консоль и в текстовый файл для каждого сезона среднее количество участвовавших в отдельной игре спортсменов, для заданного пользователем года определите имя и команду самого молодого и старого мужчины, самой высокой и низкой женщины.

**Задача 19.** В файле `data_stud_mark.txt` содержится информация о тестировании учащихся по трём предметам: пол, группа, уровень образования, условия оплаты обучения (`standard` — обычные, `free/reduced` — бесплатно или со скидкой), законченное (`completed`) или нет (`none`), баллы, полученные по трём предметам (от 0 до 100). Каждый из восьми параметров расположен в отдельной строке, в следующих восьми строках — информация о следующем объекте.

Определите для каждого предмета высший полученный балл и количество учащихся, получивших этот балл и выведите на консоль и в текстовый файл список учащихся по каждому предмету с указанием пола и группы. Определите и выведите на консоль количество женщин и мужчин, участвовавших в тестировании и их средний балл по каждому предмету и общий средний балл. Определите учащегося, набравшего самое большое количество баллов по всем трём предметам и выведите на консоль название его группы и количество тестируемых учащихся в этой группе.

Дополнительно выведите на консоль и в текстовый файл информацию о каждой группе в виде таблицы: количество женщин, количество мужчин, общий средний балл у мужчин и у женщин, максимальную и минимальную сумму баллов по всем предметам у мужчин и у женщин. По группе, предмету и граничной сумме баллов, введенных пользователем, определите и выведите на консоль долю женщин из данной группы балл которых по заданному предмету оказался выше границы.

**Задача 20.** В файле `data_weather.txt` содержится информация о погоде каждого дня 2009 — 2016 гг.: дата наблюдения, название метеостанции, минимальная температура в градусах Цельсия, максимальная температура в градусах Цельсия, количество осадков, зафиксированных за день в мм., так называемое испарение на сковороде за 24 часа, количество часов яркого солнечного света в день, направление сильнейшего порыва ветра, скорость (км/ч) самого сильного порыва ветра, направление ветра в 9 утра (NA — значение параметра не определено). Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл для каждого года даты с самой большой и самой маленькой суточной амплитудой (разность между самой высокой и самой низкой температурами воздуха за сутки), годовую амплитуду (разница между среднемесячными температурами самого тёплого и самого холодного месяцев). По заданному пользователем направлению ветра, выведите список из 10 дат с самыми сильными порывами ветра в этом направлении.

Дополнительно для заданного года выведите на консоль и в текстовый файл таблицу, где для каждой недели года вычислены: средняя температура, максимальная суточная амплитуда, среднее количество осадков, день с самым большим количеством осадков, день с самым маленьким количеством осадков.

**Задача 21.** В файле `data_weather_australia.txt` содержится информация о погоде первых восьми месяцев 2009 года в Австралии: дата наблюдения, название метеостанции, минимальная температура в градусах Цельсия, максимальная температура в градусах Цельсия, количество осадков, зафиксированных за день в мм., так называемое испарение на сковороде за 24 часа, количество часов яркого солнечного света в день, направление сильнейшего порыва ветра, скорость (км/ч)

самого сильного порыва ветра, направление ветра в 9 утра (NA — значение параметра не определено). Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл название метеостанции и дату, когда зафиксирована самая высокая и самая низкая температура, список из заданного пользователем количества самых ветренных дней (по скорости самого сильного порыва) отсортированный сначала по дате, потом по названию метеостанции. Определите направление самого сильного порыва ветра и метеостанцию, где он был зафиксирован, и выведите на консоль количество дней с таким же направлением ветра, наблюдаемым на этой метеостанции.

Дополнительно выведите на консоль и в текстовый файл таблицы, где для каждой метеостанции и каждого месяца вычислены: максимальная среднесуточная температура, максимальная суточная амплитуда (разность между самой высокой и самой низкой температурами воздуха за сутки), суммарное количество осадков, количество дней без осадков.

**Задача 22.** В файле `data_chess_player.txt` содержится информация о лучших шахматных игроках: дата составления рейтинга, рейтинг (занимаемое место), имя (фамилия и имя разделённые запятой и пробелом), звание (g — Grand Master, wg — Woman Grand Master), страна, в которой родился, стандартный шахматный рейтинг (выше 2400 — Grand Master), количество сыгранных игр за некоторый период, год рождения. Каждый из восьми параметров расположен в отдельной строке, в следующих восьми строках — информация о следующем объекте.

Определите и выведите на консоль количество и долю женщин в списке всех игроков. Для лучшего игрока среди женщин выведите на консоль и в текстовый файл список игроков, отличающихся от неё по возрасту не более, чем на пять лет. Список упорядочите по стране рождения, а затем по фамилии игрока.

Дополнительно выведите на консоль и в текстовый файл таблицу с количеством игроков для каждой страны и года рождения. Выведите на консоль список из десяти самых молодых игроков, упорядоченных по стандартному рейтингу, при выводе информации первым должно стоять имя игрока, затем фамилия.

**Задача 23.** В файле `data_computer_book.txt` содержится информация о книгах: рейтинг от 0 до 5 (оценка качества), количество отзывов (запятая отделяет тысячи), название, краткое описание, количество страниц, вид книги (Hardcover — твёрдая обложка, Paperback — мягкая обложка, ebook — электронный формат, Kindle Edition — специальный электронный формат для некоторых устройств), цена в долларах. Каждый из семи параметров расположен в отдельной строке, в следующих семи строках — информация о следующем объекте.

Для вида книги, заданного пользователем, выведите на консоль и в текстовый файл список всех книг, отсортированных по цене и отдельно тот же список, отсортированный по рейтингу. Для всех книг, рейтинг которых выше 4, выведите на консоль списки 10% самых дешёвых и 10% с самым большим количеством отзывов.

Дополнительно организуйте поиск книг по ключевому слову, введённому пользователем, ключевое слово ищите в названии и описании. Результат поиска сортируйте по рейтингу, цене, количеству страниц, в зависимости от выбора пользователя. Отсортированный список выведите на консоль. В текстовый файл выведите часть списка из первых  $k$  объектов,  $k$  задаёт пользователь.

**Задача 24.** В файле `data_countries_world.txt` содержится информация о странах: название, описание региона, численность населения, площадь в квадратных милях (1 миля примерно 1,6 км.), плотность населения на одну квадратную милю, отношение длины побережья к площади (береговая линия), миграция в год среднее количество на 1000 человек населения, младенческая смертность на 1000 рождений, ВВП на душу населения, грамотность в процентах от всего населения. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл список стран с ВВП из заданного пользователем диапазона, список отсортируйте сначала по региону, потом по площади. Для этого списка определите среднюю численность населения и названия трёх стран с самым высоким уровнем грамотности.

Дополнительно определите страну с самым длинным побережьем и выведите на консоль список стран, отличающихся от этой страны по численности населения не более чем на 10%. Список отсортируйте по одному из параметров: площадь, ВВП, грамотность. Параметр, возрастание или убывание определяет пользователь.

**Задача 25.** В файле `data_diamond.txt` содержится информация о бриллиантах (часть параметров связаны с физическими размерами: длиной  $x$ , шириной  $y$ , высотой  $z$ ): индекс, вес в каратах, качество огранки (в порядке увеличения Fair, Good, Very Good, Premium, Ideal), цвет (буква латинского алфавита от лучшего D до худшего J), прозрачность (в порядке от лучшего к худшему: FL, IF, VVS1, VVS2, VS1, VS2, SI1, SI2, I1, I2, I3), глубина (высота алмаза делённая на его средний диаметр в процентах  $\frac{z}{(x+y)/2} \times 100 = \frac{200z}{x+y}$ ), ширина алмаза в процентах от его среднего диаметра, цена бриллианта, длина  $x$  в мм, ширина  $y$  в мм. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Определите качество огранки самого дорогого бриллианта. Выведите на консоль и в текстовый файл информацию о бриллиантах этой огранки каждого цвета: количество, среднюю глубину, средний диаметр, минимальный, максимальный, средний вес.

Дополнительно, для заданной пользователем прозрачности, определите наиболее часто встречающийся цвет и выведите на консоль и в текстовый файл список бриллиантов этого цвета, отсортированный по возрастанию цены.

**Задача 26.** В файле `data_indian_dishes.txt` содержится информация об индийских блюдах: название, список ингредиентов, разделённых запятыми и пробелами, вегетарианское или нет, время подготовки продуктов, время приготовления блюда, вкус (острый, сладкий, горький), вид (основное, десерт, закуска), индийский штат — родина блюда, сторона света, к которой относится штат (север, юг, восток, запад). Каждый из девяти параметров расположен в отдельной строке, в следующих девяти строках — информация о следующем объекте. Если значение какого-то параметра неизвестно, в строке стоит -1.

Выведите на консоль и в текстовый файл для основных блюд: количество острых и не острых, для десертов: количество сладких и не сладких, для закусок: общее количество, количество острых, количество горьких. Отдельно для вегетарианских и не вегетарианских блюд выведите на консоль и в текстовый файл список из десяти блюд самых быстрых по суммарному времени готовки.

Дополнительно выведите на консоль и в текстовый файл список штатов, отсортированных по возрастанию среднего времени приготовления блюд.

## Оценка 4

Для получения оценки 4 задача должна быть решена полностью.

**Задача 27.** В файле `data_auto.txt` содержится информация о продаваемых автомобилях: индекс, цена автомобиля в объявлении, марка автомобиля, модель автомобиля, год выпуска автомобиля, техническое состояние автомобиля (clean vehicle — хорошее, salvage insurance — отремонтированный после серьёзной поломки), количество миль, пройденных автомобилем, цвет автомобиля, идентификационный номер автомобиля (набор из 17 символов — цифр и заглавных букв), номер партии выпущенных автомобилей. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Для каждого встречающегося в файле года выпуска автомобилей, выведите на консоль и в текстовый файл информацию о моделях автомобилей в виде таблицы со столбцами: средняя цена, минимальная цена, максимальная цена, самый часто встречающийся цвет автомобилей, количество автомобилей в состоянии clean vehicle, количество автомобилей в состоянии salvage insurance, средний пробег в километрах.

**Задача 28.** В файле `data_book.txt` содержится информация о книгах: индекс, название, автор книги (несколько авторов разделяются символом /), средний рейтинг книги от 0 до 5, уникальный идентификационный номер, международный стандартный номер книги ISBN, код языка, количество страниц, общее количество оценок, полученных книгой, общее количество рецензий на текст книги. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл таблицу, где для каждого автора укажите общее количество книг, среднюю толщину книги, средний рейтинг, название и ISBN самой рейтинговой книги. Для автора, выбранного пользователем, выведите на консоль список книг, отсортированный по количеству рецензий. Организуйте поиск книги по названию или ISBN.

**Задача 29.** В файле `data_chocolate.txt` содержится информация о сортах тёмного шоколада и их экспертной оценке: страна экспортер какао-бобов, название фирмы экспортера, название марки

шоколада, индекс экспертизы, год экспертизы, процентное содержание какао, страна, в которой выпускается сам шоколад указанной марки, экспертная оценка. Каждый из восьми параметров расположен в отдельной строке, в следующих восьми строках - информация о следующем объекте.

Для всех возможных значений процентного содержания какао выведите на консоль и в текстовый файл таблицу, содержащую количество выпускаемых марок шоколада, среднюю экспертную оценку, страну, выпускающую марку с лучшей экспертной оценкой, значение этой оценки, название соответствующей марки. Для выбранной пользователем страны производителя шоколада выведите на консоль список стран-экспортёров какао-бобов, отсортированных по убыванию средней экспертной оценки соответствующих марок шоколада.

**Задача 30.** В файле `data_game.txt` содержится информация о компьютерных играх: идентификатор, название игры, платформа, год выпуска, жанр, организация-разработчик, цена в Северной Америке, цена в Европе, цена в Японии, цена в других странах. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл информацию в виде таблиц. Для каждого года и каждой платформы: количество игр, для каждого жанра и каждой платформы: количество игр, для каждой организации и для каждого года: количество и среднюю цену игр, для каждой платформы: названия 10 самых дорогих игр в заданном году и заданном регионе.

**Задача 31.** В файле `data_olimp.txt` содержится информация об олимпийских спортсменах: индекс, имя, пол (М — мужской или F — женский), возраст — целое число, высота в сантиметрах, вес в килограммах (если данные не известны — NA), название команды, трехбуквенный код Национального олимпийского комитета, год и сезон олимпийских игр, год. Каждый из девяти параметров расположен в отдельной строке, в следующих девяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл информацию в виде таблиц. Для каждого года и каждого Национального олимпийского комитета: количество игроков, для каждой команды: средний возраст, средний рост, средний вес, доля мужчин и женщин. Для каждого года по возрастанию: сезон, количество мужчин и женщин, количество команд.

**Задача 32.** В файле `data_stud_mark.txt` содержится информация о тестировании учащихся по трём предметам: пол, группа, уровень образования, условия оплаты обучения (`standard` — обычные, `free/reduced` — бесплатно или со скидкой), законченное (`completed`) или нет (`none`), баллы, полученные по трём предметам (от 0 до 100). Каждый из восьми параметров расположен в отдельной строке, в следующих восьми строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл информацию в виде таблиц. Для каждой группы и уровня образования: количество протестированных учащихся, их средний балл, по выбору пользователем предмета и пола: максимальный и минимальный балл учащихся. По заданному пользователем уровню образования определите и выведите на экран пять лучших по сумме баллов результатов среди закончивших этот уровень и среди незакончивших.

**Задача 33.** В файле `data_weather.txt` содержится информация о погоде каждого дня 2009 — 2016 гг.: дата наблюдения, название метеостанции, минимальная температура в градусах Цельсия, максимальная температура в градусах Цельсия, количество осадков, зафиксированных за день в мм., так называемое испарение на сковороде за 24 часа, количество часов яркого солнечного света в день, направление сильнейшего порыва ветра, скорость (км/ч) самого сильного порыва ветра, направление ветра в 9 утра (NA — значение параметра не определено). Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл информацию в виде таблиц. Для каждого года и каждого месяца: среднюю, максимальную среднюю за день, минимальную среднюю за день температуру, среднее количество осадков, наиболее частое направление ветра. Для каждого года: количество дней с температурой выше среднегодовой, количество дней без осадков, количество дней с порывами ветра выше 25 мс. По заданному пользователем месяцу определите и выведите на консоль самое частое и самое редкое направление ветра за весь период наблюдения.

**Задача 34.** В файле `data_weather_australia.txt` содержится информация о погоде первых восьми месяцев 2009 года в Австралии: дата наблюдения, название метеостанции, минимальная температура в градусах Цельсия, максимальная температура в градусах Цельсия, количество осадков, зафиксированных за день в мм., так называемое испарение на сковороде за 24 часа, количество часов яркого солнечного света в день, направление сильнейшего порыва ветра, скорость (км/ч) самого сильного порыва ветра, направление ветра в 9 утра (NA — значение параметра не определено). Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках



— информация о следующем объекте.

Выведите на консоль и в текстовый файл по заданному пользователем месяцу сравнительную таблицу, где для каждой метеостанции и каждого дня месяца вычислены: среднесуточная температура, количество осадков, сила дождя (сильным — называется дождь, если его выпадает от 15 до 49 мм, очень сильным, когда их выпадает от 50), направление и скорость в мс самого сильного порыва ветра. Выведите на консоль для каждого месяца название метеостанции где наблюдалась самая высокая температура, самая низкая температура, наибольшее количество дней с осадками, наименьшее суммарное количество осадков.

**Задача 35.** В файле `data_chess_player.txt` содержится информация о лучших шахматных игроках: дата составления рейтинга, рейтинг (занимаемое место), имя (фамилия и имя разделённые запятой и пробелом), звание (`g` — Grand Master, `wg` — Woman Grand Master), страна, в которой родился, стандартный шахматный рейтинг (выше 2400 — Grand Master), количество сыгранных игр за некоторый период, год рождения. Каждый из восьми параметров расположен в отдельной строке, в следующих восьми строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл список стран, упорядоченный по убыванию количества игроков этой страны, попавших в рейтинг. Для текущего года выведите на консоль и в текстовый файл списки из заданного пользователем количества самых молодых и самых старших игроков. Для страны с самым большим количеством игроков выведите на консоль таблицу: в первом столбце список этих игроков, упорядоченный по составленному рейтингу, во втором столбце тот же список, упорядоченный по стандартному шахматному рейтингу.

**Задача 36.** В файле `data_computer_book.txt` содержится информация о книгах: рейтинг от 0 до 5 (оценка качества), количество отзывов (запятая отделяет тысячи), название, краткое описание, количество страниц, вид книги (`Hardcover` — твёрдая обложка, `Paperback` — мягкая обложка, `ebook` — электронный формат, `Kindle Edition` — специальный электронный формат для некоторых устройств), цена в долларах. Каждый из семи параметров расположен в отдельной строке, в следующих семи строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл информацию в виде таблиц. Для каждого вида книг и диапазона оценок (0-1, 1-2, 2-3, 3-4, 4-5): количество книг, среднюю цену, среднее количество отзывов. Сформируйте список книг, которые есть в разных видах, выведите на консоль список книг с указанием этих видов или сообщение, что таких не обнаружено.

**Задача 37.** В файле `data_countries_world.txt` содержится информация о странах: название, описание региона, численность населения, площадь в квадратных милях (1 миля примерно 1,6 км.), плотность населения на одну квадратную милю, отношение длины побережья к площади (береговая линия), миграция в год среднее количество на 1000 человек населения, младенческая смертность на 1000 рождений, ВВП на душу населения, грамотность в процентах от всего населения. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Для каждого региона выведите на консоль и в текстовый файл информацию в виде таблицы: количество стран, общая численность населения, общая площадь, количество стран без выхода к морю или океану, самое большое количество человек уехавших из некоторой страны (на основе численности населения и уровня миграции), среднюю грамотность 5 самых богатых по уровню ВВП стран, страну с самой низкой детской смертностью. Регионы отсортируйте по общей численности населения.

**Задача 38.** В файле `data_diamond.txt` содержится информация о бриллиантах (часть параметров связаны с физическими размерами: длиной  $x$ , шириной  $y$ , высотой  $z$ ): индекс, вес в каратах, качество огранки (в порядке увеличения `Fair`, `Good`, `Very Good`, `Premium`, `Ideal`), цвет (буква латинского алфавита от лучшего `D` до худшего `J`), прозрачность (в порядке от лучшего к худшему: `FL`, `IF`, `VVS1`, `VVS2`, `VS1`, `VS2`, `SI1`, `SI2`, `I1`, `I2`, `I3`), глубина (высота алмаза делённая на его средний диаметр в процентах  $\frac{z}{(x+y)/2} \times 100 = \frac{200z}{x+y}$ ), ширина алмаза в процентах от его среднего диаметра, цена бриллианта, длина  $x$  в мм, ширина  $y$  в мм. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл информацию в виде таблиц для каждого качества огранки и цвета: максимальную, минимальную, среднюю цену, средний диаметр, наиболее часто встречающуюся прозрачность (если наибольшее количество объектов соответствует нескольким видам прозрачности, выберите лучшую).

**Задача 39.** В файле `data_indian_dishes.txt` содержится информация об индийских блюдах: на-

звание, список ингредиентов, разделённых запятыми и пробелами, вегетарианское или нет, время подготовки продуктов, время приготовления блюда, вкус (острый, сладкий, горький), вид (основное, десерт, закуска), индийский штат — родина блюда, сторона света, к которой относится штат (север, юг, восток, запад). Каждый из девяти параметров расположен в отдельной строке, в следующих девяти строках — информация о следующем объекте. Если значение какого-то параметра неизвестно, в строке стоит -1.

Выведите на консоль и в текстовый файл информацию в виде таблиц для каждого штата и вида блюда: общее количество блюд, долю не вегетарианских блюд, среднее время подготовки продуктов, среднее время приготовления самого блюда, максимальное и минимальное суммарное время приготовления, долю острых блюд. Отдельно выведите список штатов, упорядоченный по убыванию доли острых блюд среди всех.

**Задача 40.** В файле `data_meteorite.txt` содержится информация о метеоритах: название, идентификатор, тип (`valid` — обычный, `relict` — метеорит, сильно разрушенный погодными условиями), класс (классификация зависит от физических, химических и некоторых других свойств метеорита), масса в граммах, наблюдалось ли падение (`Fell` — наблюдалось падение метеорита, `Found` — падения метеорита не наблюдалось), год обнаружения метеорита, широта падения метеорита, долгота падения метеорита, геолокация в виде точки с парой координат: широтой и долготой. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл информацию в виде таблицы для каждого века и типа количество обнаруженных метеоритов. Отдельно для каждого десятилетия 19 и 20 веков выведите гистограмму количества обнаруженных метеоритов. Определите год, десятилетие, двадцатилетие, когда наблюдалось максимальное падение метеоритов, для каждого найденного периода выведите на консоль информацию о периоде и список всех обнаруженных в этот период метеоритов (не только наблюдаемых), упорядоченных по широте.

## Оценка 4 — 5

Для получения оценки 4 нужно решить всё, кроме дополнительного задания. Для получения оценки 5 задача должна быть решена полностью.

**Задача 41.** В файле `data_anime.txt` содержится информация о фильмах и сериалах в стиле аниме: уникальный идентификатор, полное название аниме, жанр (список жанров этого аниме, разделённых запятыми), тип, количество серий (1, если фильм), рейтинг - средняя оценка этого аниме от 0 до 10, количество участников сообщества в этом аниме. Каждый из семи параметров расположен в отдельной строке, в следующих семи строках - информация о следующем объекте.

Выведите на консоль таблицу, где для каждого жанра аниме укажите общее количество аниме этого жанра, средний рейтинг, максимальный рейтинг, минимальный рейтинг, среднее количество участников сообществ. По выбору пользователя выведите аналогичную таблицу для фильмов или сериалов, а также список 20 лучших аниме, выбранного пользователем вида.

Дополнительно реализуйте поиск аниме по параметрам, вводимым пользователем: жанр, тип, диапазон серий, диапазон рейтинга, диапазон количества участников сообщества. Для каждого параметра должен быть предусмотрен ввод значения «любой». Выведите результат поиска на консоль и в файл в формате `json`.

**Задача 42.** В файле `data_course.txt` содержится информация об онлайн-курсах: идентификатор курса, название курса, URL курса, курс бесплатный (`False`) или платный (`True`), стоимость курса (0, если курс бесплатный), текущее количество слушателей, количество отзывов, количество лекций, сложность курса, продолжительность курса в месяцах. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Организируйте подбор курса для пользователя по его запросу, параметры запроса: бесплатный или платный, если платный — диапазон стоимости, сложность (выбор из заданного набора), верхняя граница длительности. Если в результате количество курсов превышает 10, выберите те у которых больше слушателей или меньше количество лекций, приоритет этих критериев можно спросить у пользователя или построить суммарный критерий из значений этих двух параметров с коэффициентами. Если нет подходящих бесплатных курсов, найдите подходящие платные и выведите их отдельным списком, и наоборот, если есть бесплатные, подходящие по критериям, то сообщите о

них. Результат поиска выведите на консоль или в текстовый файл.

Дополнительно попросите пользователя выбрать из результатов поиска наиболее понравившийся курс и найдите для него близкие по параметрам. Числовые параметры не должны при этом сильно отличаться от запроса (до 10%), а название курса может содержать похожие слова (совпадающие или отличающиеся окончанием). Приоритет критериев можно установить самостоятельно, а для некоторых сделать запрос у пользователя (можно посмотреть как построить функцию полезности из нескольких критериев функцию полезности из нескольких критериев).

**Задача 43.** В файле `data_journals.txt` содержится информация о научных журналах: название, импакт-фактор (международный рейтинг), категория, количество цитирований статей журнала за последние три года, количество опубликованных статей за последние три года, процентное соотношение цитирования по сравнению с предыдущим годом, коэффициент цитирований на основе общего количества цитирований в предметной области, коэффициент цитирования в рейтинговых журналах, название издателя журнала. Каждый из девяти параметров расположен в отдельной строке, в следующих девяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл списки журналов каждого издательства, упорядоченные по импакт-фактору или одному из коэффициентов цитирования по выбору пользователя. Разделите журналы по тематике на списки: медицина, биология, физика, химия, другое, и каждый список отсортируйте по среднему количеству цитат на одну статью за последние три года. Выведите на консоль из списка, выбранного пользователем (по издательству или по тематике) заданное количество первых журналов.

Дополнительно реализуйте сортировку журналов по обобщённому критерию на основе импакт-фактора, количества статей, коэффициентов цитирования (можно посмотреть как построить функцию полезности из нескольких критериев функцию полезности из нескольких критериев). Выведите в текстовые файлы отсортированные таким образом список всех журналов и списки по издательствам и тематике.

**Задача 44.** В файле `data_olimp.txt` содержится информация об олимпийских спортсменах: индекс, имя, пол (М — мужской или F — женский), возраст — целое число, высота в сантиметрах, вес в килограммах (если данные не известны — NA), название команды, трехбуквенный код Национального олимпийского комитета, год и сезон олимпийских игр, год. Каждый из девяти параметров расположен в отдельной строке, в следующих девяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл информацию в виде таблицы с количеством спортсменов для каждого года, сезона и команды, а также информацию об участии команд каждого Национального олимпийского комитета в каждой игре.

Дополнительно найдите список спортсменов заданной команды в заданном году, упорядоченный по полу, а затем по возрасту. Для каждой команды вычислите среднее количество участников в год, средний, минимальный, максимальный возраст, средний рост и вес, максимальную и минимальную долю мужчин. Выведите результат поиска на консоль и в файл в формате `json`.

**Задача 45.** В файле `data_stud_mark.txt` содержится информация о тестировании учащихся по трём предметам: пол, группа, уровень образования, условия оплаты обучения (`standard` — стандартная оплата, `free/reduced` — бесплатно или со скидкой), законченное (`completed`) или нет (`none`), баллы, полученные по трём предметам (от 0 до 100). Каждый из восьми параметров расположен в отдельной строке, в следующих восьми строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл информацию в виде таблицы учащихся с баллами по трём предметам. Учащиеся отбираются по параметрам, задаваемым пользователем: пол, группа, уровень образования, условия оплаты, законченное или нет, диапазон полученных баллов по одному из предметов. Пользователь может задавать не все параметры, а только часть. Строки таблицы сортируются по возрастанию общей суммы баллов. Для каждого столбца таблицы считается и выводится среднее значение баллов.

Дополнительно выведите на консоль и в текстовый файл гистограммы распределения баллов по заданному пользователем предмету и уровню обучения для тех кто учится платно и тех, кто учится бесплатно. Для каждого уровня образования определите количество всех учащихся, долю учащихся, закончивших этот уровень, их общий средний балл, общий средний балл среди всех учащихся этого уровня, названия групп, в которых был получен максимальный балл по каждому из трёх предметов. Выведите результат поиска на консоль и в файл в формате `json`.

**Задача 46.** В файле `data_weather.txt` содержится информация о погоде каждого дня 2009 — 2016 гг.: дата наблюдения, название метеостанции, минимальная температура в градусах Цельсия,

максимальная температура в градусах Цельсия, количество осадков, зафиксированных за день в мм., так называемое испарение на сковороде за 24 часа, количество часов яркого солнечного света в день, направление сильнейшего порыва ветра, скорость (км/ч) самого сильного порыва ветра, направление ветра в 9 утра (NA — значение параметра не определено). Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль для каждого года три дня с самым большим количеством осадков и информацию об осадках и ветре в эти дни. По заданной пользователем дате выведите на консоль и в текстовый файл сравнительную таблицу погоды в этот день в каждом году: минимальную и максимальную температуру, осадки (сильным — называется дождь, если его выпадает от 15 до 49 мм, очень сильным, когда их выпадает от 50, сильный снег — количество осадков от 7 до 19 мм, очень сильный снег, когда его количество превышает 20), направление ветра, степень самого сильного порыва ветра (ветер скоростью порядка 0-5 м/с считается слабым, 6-14 м/с — умеренным; свыше 14 м/с — сильным; с 25 м/с — очень сильным, а выше 33 м/с — ураганом).

Дополнительно определите и выведите на консоль месяц с максимальной и минимальной средней температурой, месяц с максимальным и минимальным количеством осадков за весь период наблюдения. Для каждого сезона (зима, весна, лето, осень) найдите минимальное, максимальное, среднее значение температуры, наиболее частое направление ветра, максимальную скорость ветра. Выведите результат поиска на консоль и в файл в формате json.

**Задача 47.** В файле `data_computer_book.txt` содержится информация о книгах: рейтинг от 0 до 5 (оценка качества), количество отзывов (запятая отделяет тысячи), название, краткое описание, количество страниц, вид книги (Hardcover — твёрдая обложка, Paperback — мягкая обложка, ebook — электронный формат, Kindle Edition — специальный электронный формат для некоторых устройств), цена в долларах. Каждый из семи параметров расположен в отдельной строке, в следующих семи строках — информация о следующем объекте.

Организируйте поиск книги по параметрам, заданным пользователем: название, вид книги, диапазон цены. Результат поиска отсортируйте по отношению качество/цена и выведите список на консоль и в текстовый файл.

Дополнительно найдите для книги в описании упоминание имени автора (два слова подряд с большой буквы: имя и фамилия) или авторов, выведите список книг, для которых авторы нашлись, на консоль и в текстовый файл, упорядочив по фамилиям авторов. Оцените вручную, сколько произошло ошибок.

**Задача 48.** В файле `data_countries_world.txt` содержится информация о странах: название, описание региона, численность населения, площадь в квадратных милях (1 миля примерно 1,6 км.), плотность населения на одну квадратную милю, отношение длины побережья к площади (береговая линия), миграция в год среднее количество на 1000 человек населения, младенческая смертность на 1000 рождений, ВВП на душу населения, грамотность в процентах от всего населения. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл таблицу количества стран по плотности населения (диапазоны в количестве человек на квадратный километр: 0–25, 25–75, 75–150, 150–300, 300–1000, больше 1000), таблицу количества стран по уровню грамотности (диапазоны в процентах: 0–20, 20–40, 40–60, 60–80, 80–90, 90–100), список из десяти самых богатых по уровню ВВП стран и десяти самых бедных, для каждой страны в этих списках рассчитайте количество приезжающих или уезжающих на основе численности населения и уровня миграции и количество грамотных людей.

Дополнительно выведите указанную выше информацию отдельно для каждой части света. Часть света определите на основе описания региона, например, в полуавтоматическом режиме: напишите дополнительный программный проект, в котором выводите описание региона очередной страны и вручную указывайте к какой части света относится этот регион, в дальнейшем уже без обращения к пользователю по совпадающему описанию сразу определяйте часть света. Также можно добавить поиск соответствующих слов в описании.

**Задача 49.** В файле `data_diamond.txt` содержится информация о бриллиантах (часть параметров связаны с физическими размерами: длиной  $x$ , шириной  $y$ , высотой  $z$ ): индекс, вес в каратах, качество огранки (в порядке увеличения Fair, Good, Very Good, Premium, Ideal), цвет (буква латинского алфавита от лучшего D до худшего J), прозрачность (в порядке от лучшего к худшему: FL, IF, VVS1, VVS2, VS1, VS2, SI1, SI2, I1, I2, I3), глубина (высота алмаза делённая на его средний диаметр в процентах  $\frac{z}{(x+y)/2} \times 100 = \frac{200z}{x+y}$ ), ширина алмаза в процентах от его среднего диаметра, цена

бриллианта, длина  $x$  в мм, ширина  $y$  в мм. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Организуйте подбор бриллиантов по запросу пользователя, параметры запроса: диапазон веса, качество огранки, диапазон цены, цвет. Если по результатам подбора в списке больше 10 объектов, предложите дополнительно выбрать прозрачность из тех, что гарантированно есть в списке, если объектов по прежнему много, то выберите десять самых крупных по среднему диаметру. Результат выведите на консоль и в текстовый файл.

Дополнительно если нет объектов, подходящих пользователю по запросу, сформируйте список с объектами близкими по параметрам, числовые не должны отличаться более, чем на 5%, качественные на одну градацию лучше или хуже. Приоритет параметров запросите у пользователя. Выведите результат подбора в файл в формате json.

**Задача 50.** В файле `data_indian_dishes.txt` содержится информация об индийских блюдах: название, список ингредиентов, разделённых запятыми и пробелами, вегетарианское или нет, время подготовки продуктов, время приготовления блюда, вкус (острый, сладкий, горький), вид (основное, десерт, закуска), индийский штат — родина блюда, сторона света, к которой относится штат (север, юг, восток, запад). Каждый из девяти параметров расположен в отдельной строке, в следующих девяти строках — информация о следующем объекте. Если значение какого-то параметра неизвестно, в строке стоит -1.

Организуйте выбор блюда для пользователя по следующим параметрам: штат или сторона света, вид, отсутствие заданного ингредиента. Полученный список блюд отсортируйте по времени приготовления самого блюда. Отдельно выберите из полученного два самых сложных и два самых простых блюда. Считайте, что блюдо сложнее, если у него больше ингредиентов, а затем больше суммарное время подготовки и приготовления. Результат выведите на консоль и в текстовый файл.

Дополнительно для каждого штата и вида блюда определите наиболее часто встречающийся ингредиент, результат выведите в виде таблицы на консоль и в текстовый файл.

**Задача 51.** В файле `data_meteorite.txt` содержится информация о метеоритах: название, идентификатор, тип (`valid` — обычный, `relict` — метеорит, сильно разрушенный погодными условиями), класс (классификация зависит от физических, химических и некоторых других свойств метеорита), масса в граммах, наблюдалось ли падение (`Fell` — наблюдалось падение метеорита, `Found` — падения метеорита не наблюдалось), год обнаружения метеорита, широта падения метеорита, долгота падения метеорита, геолокация в виде точки с парой координат: широтой и долготой. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль и в текстовый файл в виде таблицы для каждого класса метеоритов: общее количество обнаруженных метеоритов, долю обычных (не разрушенных), среднюю массу всех метеоритов, название, год обнаружения и массу самого тяжёлого метеорита. Список классов отсортируйте по убыванию доли метеоритов, у которых наблюдалось падение.

Дополнительно по точке геолокации и радиусе в километрах, заданных пользователем, найдите список метеоритов, которые попали в заданную область с центром в указанной точке и соответствующим радиусом, список отсортируйте по возрастанию года обнаружения и массе. Результат выведите на консоль и в файл в формате json.

## Оценка 5

Для получения оценки 5 задача должна быть решена полностью.

**Задача 52.** В файле `data_auto.txt` содержится информация о продаваемых автомобилях: индекс, цена автомобиля в объявлении, марка автомобиля, модель автомобиля, год выпуска автомобиля, техническое состояние автомобиля (`clean vehicle` — хорошее, `salvage insurance` — отремонтированный после серьёзной поломки), количество миль, пройденных автомобилем, цвет автомобиля, идентификационный номер автомобиля (набор из 17 символов — цифр и заглавных букв), номер партии выпущенных автомобилей. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Проверьте уникальность идентификационных номеров в данных файла. Найдите автомобили из одной и той же партии и выведите соответствующие списки или сообщение, о том, что такие не

найжены на консоль и в текстовый файл. Выведите на консоль и в текстовый файл список марок автомобилей, упорядоченный по возрастанию средней цены автомобилей этих марок.

Для каждого цвета, встречающегося в файле постройте в консоли гистограмму количества автомобилей разных марок, если для какой-то из марок нет автомобилей текущего цвета, такая марка в соответствующую диаграмму не включается. Для каждой марки автомобилей выведите таблицу с информацией о средней, минимальной, максимальной цене и пробегу по годам выпуска.

Реализуйте подбор автомобиля по параметрам, вводимым пользователем: возраст или диапазон лет, цвет или несколько вариантов цвета, марка автомобиля, техническое состояние. Результат подбора выведите на консоль и в текстовый файл в виде списка, упорядоченного по выбору пользователя: по возрастанию цены, убыванию цены, возрастанию пробега, убыванию пробега.

**Задача 53.** В файле `data_chocolate.txt` содержится информация о сортах тёмного шоколада и их экспертной оценке: страна экспортер какао-бобов, название фирмы экспортера, название марки шоколада, индекс экспертизы, год экспертизы, процентное содержание какао, страна, в которой выпускается сам шоколад указанной марки, экспертная оценка. Каждый из восьми параметров расположен в отдельной строке, в следующих восьми строках - информация о следующем объекте.

Определите количество стран-экспортеров какао-бобов для каждой части света: азии, африки, америки, австралии, европы. В консоли постройте гистограмму на основе процентного соотношения стран-экспортеров по частям света. Для каждой части света постройте рейтинг стран-экспортеров какао-бобов на основе средней экспертной оценки, проведенной в заданном пользователем году, на консоль выведите 10 лучших стран-экспортеров для каждой части света.

В текстовый файл выведите таблицу для стран, выпускающих шоколад, в которой укажите количество марок шоколада, количество стран экспортеров, минимальное и максимальное процентное содержание какао, название лучшей марки шоколада и год соответствующей экспертной оценки.

**Задача 54.** В файле `data_olimp.txt` содержится информация об олимпийских спортсменах: индекс, имя, пол (М — мужской или F — женский), возраст — целое число, высота в сантиметрах, вес в килограммах (если данные не известны — NA), название команды, трехбуквенный код Национального олимпийского комитета, год и сезон олимпийских игр, год. Каждый из девяти параметров расположен в отдельной строке, в следующих девяти строках — информация о следующем объекте.

Выведите на консоль список Национальных олимпийских комитетов, у которых в играх участвовало больше одной команды. Выведите на консоль список игр (год и сезон), упорядоченных по количеству участников, затем по количеству стран (олимпийских комитетов). По заданным пользователем параметрам: команда или олимпийский комитет, год, выведите на консоль список участников команды.

Для каждого континента найдите минимальное и максимальное количество команд, участвовавших в отдельных олимпийских играх, команду, участвовавшую чаще всего, команду с самым большим количеством игроков, имена самых молодых игроков, самых старых игроков, их возраст и название команды. Выведите результат поиска на консоль и в файл в формате json.

**Задача 55.** В файле `data_weather.txt` содержится информация о погоде каждого дня 2009 — 2016 гг.: дата наблюдения, название метеостанции, минимальная температура в градусах Цельсия, максимальная температура в градусах Цельсия, количество осадков, зафиксированных за день в мм., так называемое испарение на сковороде за 24 часа, количество часов яркого солнечного света в день, направление сильнейшего порыва ветра, скорость (км/ч) самого сильного порыва ветра, направление ветра в 9 утра (NA — значение параметра не определено). Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Выведите на консоль гистограмму — аналог розы ветров для каждого года отдельно и для всего временного периода в целом. По заданной пользователем дате определите погоду на ближайшие выходные: диапазон и среднюю температуру, осадки (сильным — называется дождь, если его выпадает от 15 до 49 мм, очень сильным, когда их выпадает от 50, сильный снег — количество осадков от 7 до 19 мм, очень сильный снег, когда его количество превышает 20), направление ветра.

Для каждого года и месяца года найдите наиболее частое направление ветра, максимальную скорость ветра и соответствующее направление. Выведите результат поиска на консоль и в файл в формате json.

**Задача 56.** В файле `data_weather_australia.txt` содержится информация о погоде первых восьми месяцев 2009 года в Австралии: дата наблюдения, название метеостанции, минимальная температура в градусах Цельсия, максимальная температура в градусах Цельсия, количество осадков, зафиксированных за день в мм., так называемое испарение на сковороде за 24 часа, количество

часов яркого солнечного света в день, направление сильнейшего порыва ветра, скорость (км/ч) самого сильного порыва ветра, направление ветра в 9 утра (NA — значение параметра не определено). Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Для каждой метеостанции выведите на консоль и в текстовый файл в формате json информацию о периодах повышения и понижения среднесуточной температуры. Для каждого такого периода длиной не менее трёх дней определите максимальную суточную амплитуду (разность между самой высокой и самой низкой температурами воздуха за сутки). Для каждой метеостанции по заданному пользователем периоду времени составьте сравнительную таблицу направления ветра в 9 утра ежедневно. Определите даты, в которых направление ветра совпадает и сравните в эти дни скорость и направление самого сильного порыва ветра (укажите значения и сообщение о совпадении). Для заданной пользователем метеостанции определите даты резкой смены направления ветра в 9 утра и самого сильного порыва, считается, что происходит резкая смена ветра, если угол между направлениями ветра больше 120 градусов. Для той же станции определите даты изменения максимальной температуры выше порога, заданного пользователем, и выведите количество осадков в эти даты.

**Задача 57.** В файле `data_computer_book.txt` содержится информация о книгах: рейтинг от 0 до 5 (оценка качества), количество отзывов (запятая отделяет тысячи), название, краткое описание, количество страниц, вид книги (Hardcover — твёрдая обложка, Paperback — мягкая обложка, ebook — электронный формат, Kindle Edition — специальный электронный формат для некоторых устройств), цена в долларах. Каждый из семи параметров расположен в отдельной строке, в следующих семи строках — информация о следующем объекте.

Для каждого вида книги выведите на консоль гистограмму распределения количества отзывов по тысячам, выведите в файл в формате json средний рейтинг, список из пяти названий самых рейтинговых книг, среднюю цену, диапазон цен, диапазон страниц, список названий книг, для которых нет ни одного отзыва.

Для выбранной пользователем книги найдите наиболее похожие от одной до пяти. Чем больше одинаковых слов в названии и описании, тем более похожими считайте книги.

**Задача 58.** В файле `data_countries_world.txt` содержится информация о странах: название, описание региона, численность населения, площадь в квадратных милях (1 миля примерно 1,6 км.), плотность населения на одну квадратную милю, отношение длины побережья к площади (береговая линия), миграция в год среднее количество на 1000 человек населения, младенческая смертность на 1000 рождений, ВВП на душу населения, грамотность в процентах от всего населения. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Проведите исследование степени связи между всеми возможными парами следующих параметров: численность, площадь, миграция, младенческая смертность, ВВП, грамотность. Степень связи определите по статистическому критерию: коэффициенту корреляции (описание коэффициента корреляции, пример расчёта). Выведите на консоль и в текстовый файл список пар параметров по убыванию степени связи.

**Задача 59.** В файле `data_diamond.txt` содержится информация о бриллиантах (часть параметров связаны с физическими размерами: длиной  $x$ , шириной  $y$ , высотой  $z$ ): индекс, вес в каратах, качество огранки (в порядке увеличения Fair, Good, Very Good, Premium, Ideal), цвет (буква латинского алфавита от лучшего D до худшего J), прозрачность (в порядке от лучшего к худшему: FL, IF, VVS1, VVS2, VS1, VS2, SI1, SI2, I1, I2, I3), глубина (высота алмаза делённая на его средний диаметр в процентах  $\frac{z}{(x+y)/2} \times 100 = \frac{200z}{x+y}$ ), ширина алмаза в процентах от его среднего диаметра, цена бриллианта, длина  $x$  в мм, ширина  $y$  в мм. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Отсортируйте общий список бриллиантов сразу по четырём параметрам: качеству огранки, затем цвету, прозрачности, весу. Организуйте подбор бриллианта по его размерам: диапазону длины, ширины, высоты. Список отобранных объектов отсортируйте по указанным параметрам, порядок параметров при сортировке задаёт пользователь. Каждый запрос пользователя и результат отбора фиксируйте в файле: дата и время запроса, параметры запроса, отобранный список в виде набора индексов в соответствующем выбранной сортировке порядке.

**Задача 60.** В файле `data_indian_dishes.txt` содержится информация об индийских блюдах: название, список ингредиентов, разделённых запятыми и пробелами, вегетарианское или нет, время подготовки продуктов, время приготовления блюда, вкус (острый, сладкий, горький), вид (основ-

ное, десерт, закуска), индийский штат — родина блюда, сторона света, к которой относится штат (север, юг, восток, запад). Каждый из девяти параметров расположен в отдельной строке, в следующих девяти строках — информация о следующем объекте. Если значение какого-то параметра неизвестно, в строке стоит -1.

По заданному пользователем списку ингредиентов определите блюда, которые можно из них приготовить, если таких нет, выберите наиболее подходящие, для которых среди заданных не хватает не более одного продукта, полученный список блюд выведите на консоль и в текстовый файл, отсортировав по виду и возрастанию количества ингредиентов.

По заданному пользователем блюду выберите и выведите на консоль три наиболее похожих блюда его вида. Можно представить каждое блюдо в виде вектора из нулей и единиц по наличию ингредиентов и считать расстояние Хэмминга. Если одинаково похожих больше трёх, выберите те, которые быстрее готовятся.

**Задача 61.** В файле `data_meteorite.txt` содержится информация о метеоритах: название, идентификатор, тип (`valid` — обычный, `relict` — метеорит, сильно разрушенный погодными условиями), класс (классификация зависит от физических, химических и некоторых других свойств метеорита), масса в граммах, наблюдалось ли падение (`Fell` — наблюдалось падение метеорита, `Found` — падения метеорита не наблюдалось), год обнаружения метеорита, широта падения метеорита, долгота падения метеорита, геолокация в виде точки с парой координат: широтой и долготой. Каждый из десяти параметров расположен в отдельной строке, в следующих десяти строках — информация о следующем объекте.

Придумайте и реализуйте алгоритм определения части света по широте и долготе, можно использовать карту мира. Для каждой стороны света определите общее количество обнаруженных метеоритов и долю тех, для которых наблюдалось падение, среднюю массу метеоритов, название, год обнаружения и массу самого тяжёлого и самого лёгкого метеорита. Результат выведите в файл в формате `json`, на консоль выведите список частей света, упорядоченных по количеству обнаруженных метеоритов.