



ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ВЫСШАЯ ШКОЛА ЭКОНОМИКИ

Учебная практика Основы C++

Большой проект
2016-2017



Цель работы

Цель работы – экспериментальное определение времени сортировок различных массивов разными методами и сравнительный анализ полученных результатов



Задание

Провести эксперименты – измерить среднее время выполнения сортировок массива по возрастанию методами

- 1) пузырька
- 2) пузырька с условием Айверсона 1
- 3) пузырька с условием Айверсона 1+2
- 4) простых вставок
- 5) бинарных вставок
- 6) подсчетом (устойчивая)
- 7) цифровая

Уберите из функций сортировок все постороннее: копирование массивов, счетчики операций, вывод и т.п.



Задание. Продолжение

1. Измерения провести для размеров массива от 1000 до 9000, шаг 1000
2. Измерения выполнить для массивов, заполненных целыми неотрицательными числами:
 - случайными значениями от 0 до 7
 - случайными значениями от 0 до максимального целого (кроме подсчета)
 - «Почти» отсортированными в требуемом порядке числами (например, поменять местами 5 пар элементов отсортированного массива)
 - Отсортированными в обратном порядке (по убыванию) числами от 0 до 8000
3. Результаты измерений вывести в файл / файлы *.csv (разделители – точки с запятой «;»)



Задание. Продолжение

Каждую функцию вызывать по 100 раз, найти среднее время выполнения сортировки (**в тактах или в наносекундах**) для массивов заданных размеров.

Сортировать всякий раз один и тот же массив. Не сортировать сортированный на предыдущем шаге!!!

В цикле для определения времени сгенерировать эталонный массив максимальной длины 9000 (4 раза – для 4 видов массивов), затем перед каждой сортировкой копировать элементы эталонного массива в рабочий массив, который и сортировать.

Сначала копируете 1000 элементов, потом 2000 и т.д.



Задание. Результаты

Представить результаты экспериментов в виде таблиц в Excel, в которых указать: Метод сортировки, вид массива (например, «пузырек, случайные [0;7]» или «Простые вставки, обратно сортированный»)

Примерный вид таблицы:

	Пузырек Случайные числа 0-7
Размер массива	Среднее время сортировки (единицы измерения)
1000	
2000	

Таблицы с результатами экспериментов – на листах электронной таблицы
Для этого выводите данные в файлы *.csv с разделителем ;
Потом загрузите эти файлы в ЭТ (импорт)



ВАЖНО!!!

Первые прогоны программы выполняются медленнее, т.к. оптимизирующий компилятор еще не провел оптимизацию.

Можно оставить время как есть, но пояснить на защите проекта.

Лучше исключить 2-3 первых замера времени сортировок из расчета среднего времени



Задание. Результаты

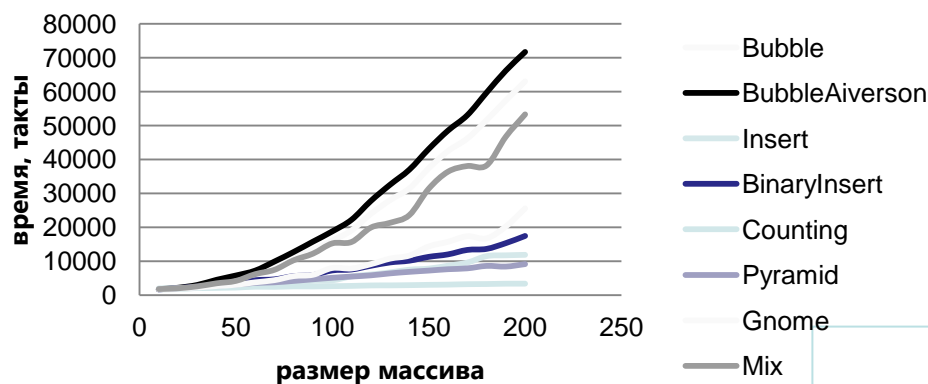
Представить результаты экспериментов в виде графиков (в ЭТ) «размер массива – среднее время сортировки».

1. По оси Y – время, указать единицы измерения (такты или наносекунды), по оси X – размер массива
2. 4 графика, для каждого вида массива. Легенды – метод сортировки (4 графика, т.к 4 вида заполнения массива, на каждом по 7 (или 6) кривых $t_{\text{среднее}}$, соответствующих семи методам сортировки
3. 7 графиков, для каждого метода сортировки. Легенды – вид массива (на каждом графике по 4 кривых $t_{\text{среднее}}$, по количеству разных массивов)



Пример графиков 1

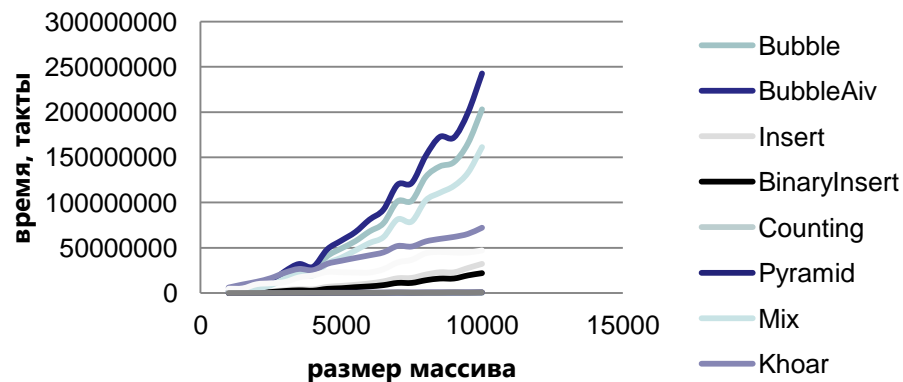
значения массива - случайные 0-5



Оси на графиках должны быть оцифрованы, обозначены с указанием единиц измерения (для времени)

Если при размещении всех методов на одном графике для одного вида исходного массива невозможно их сравнить (некоторые графики «лежат» на оси X), можно разделить методы на 2 группы и построить 2 графика)

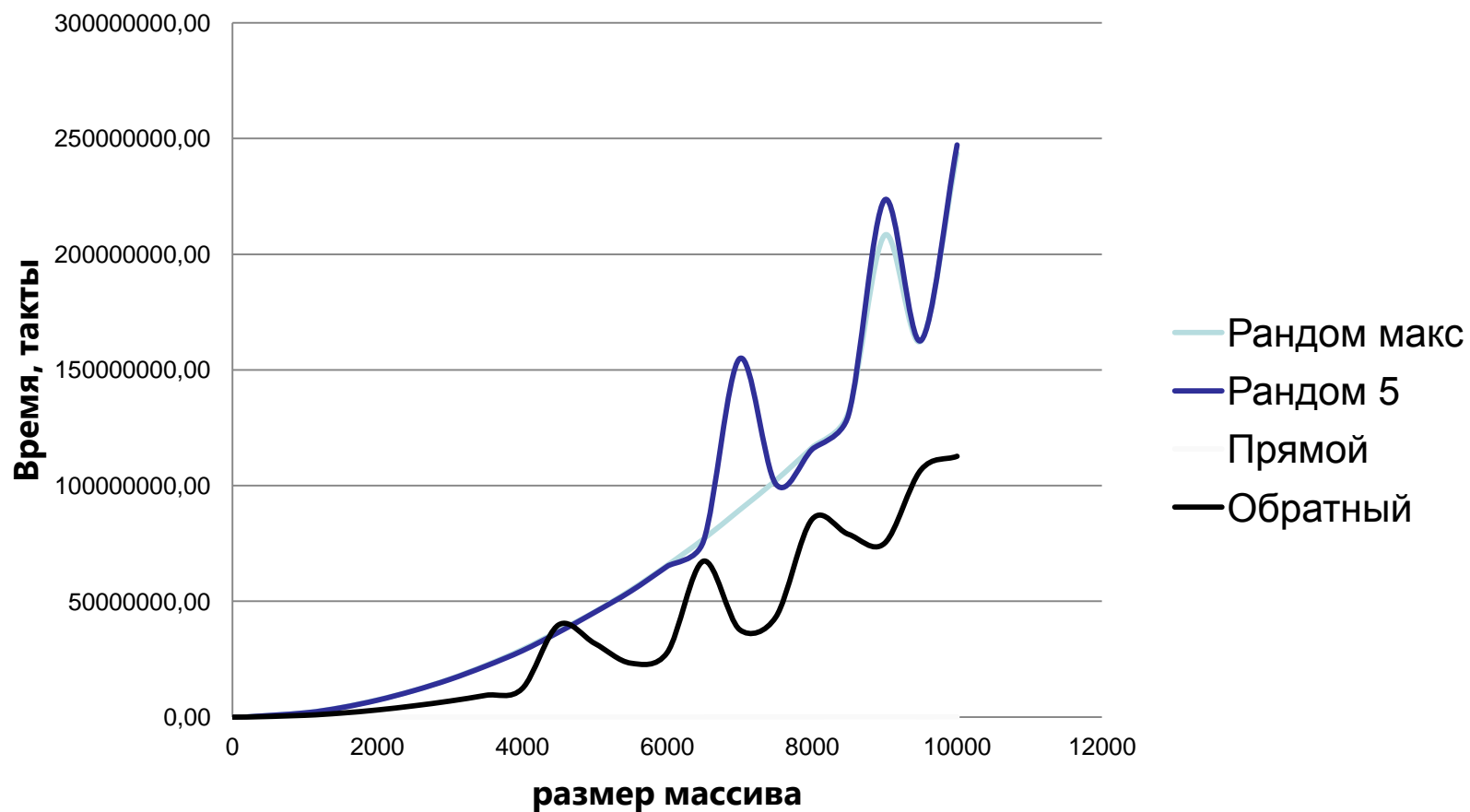
обратно сортированный массив





Пример графиков 2

Пузырек с условием Айверсона 1

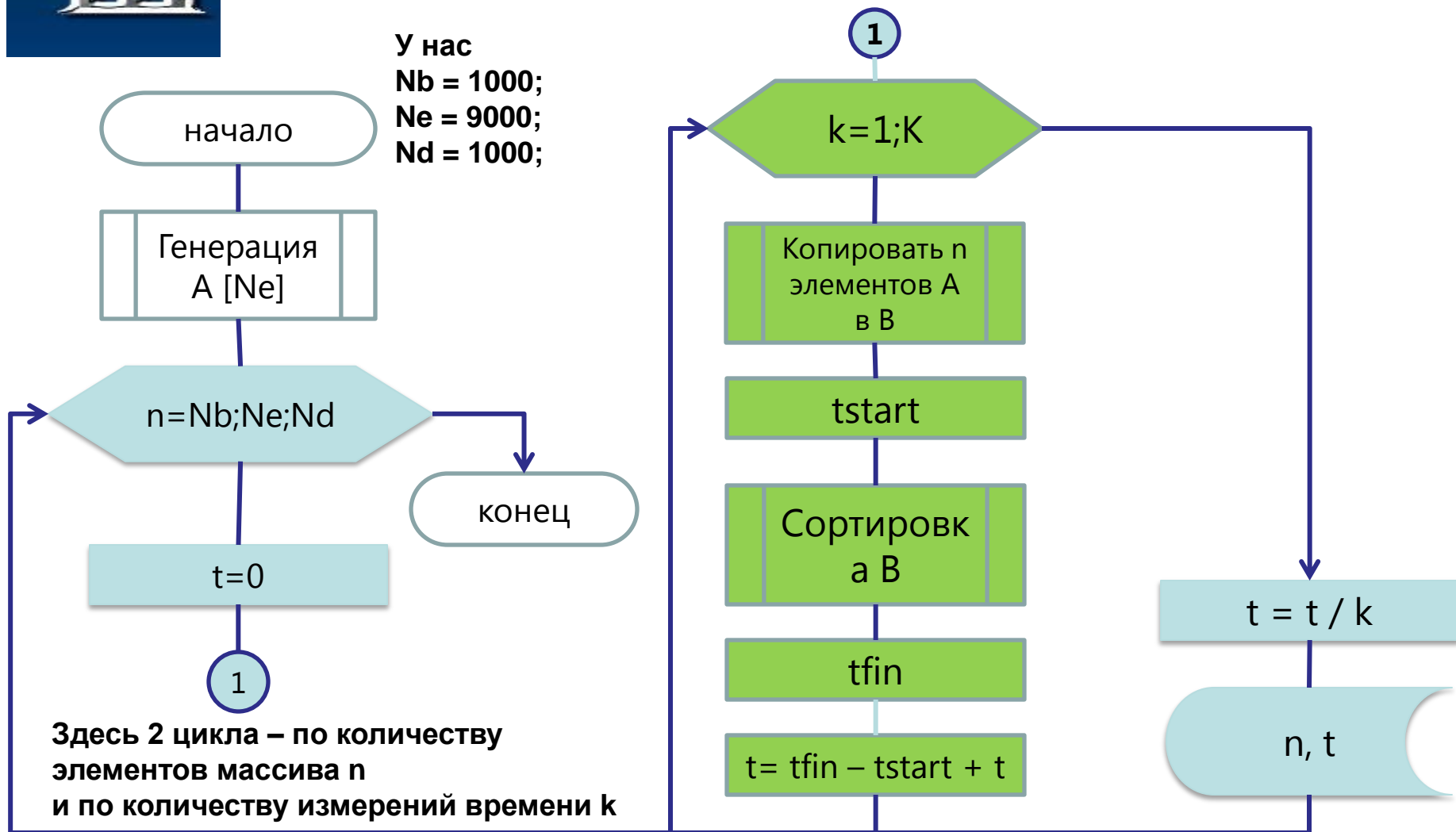




Эксперимент попроще

(один вид массива, один метод сортировки)

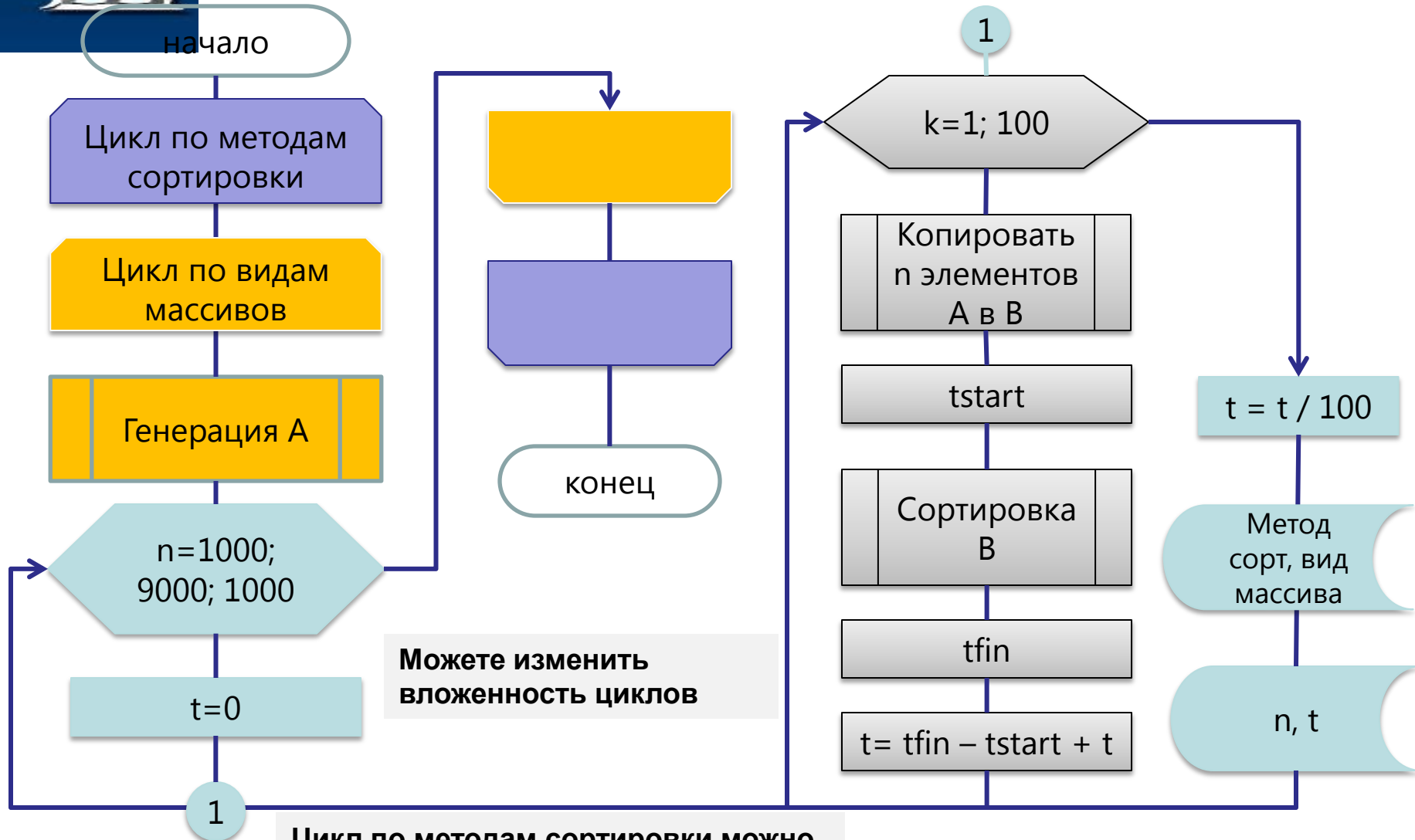
У нас
 $Nb = 1000$;
 $Ne = 9000$;
 $Nd = 1000$;





Эксперимент сложнее

Здесь – 4 вложенных цикла



Цикл по методам сортировки можно использовать, если создать массив указателей на функции



Обозначения в блок-схемах

A – эталонный массив

B – рабочий массив

K – количество повторов для вычисления среднего времени сортировки

n – текущий размер массива для сортировки

Nb – начальный размер массива

Ne – конечный размер массива

Nd – шаг изменения размера массива



Комментарии в коде

В начале кода обязательно указать

- 1) Учебная практика 2017
- 2) ФИО и группа
- 3) Среда разработки
- 4) Перечислить, что сделано
- 5) Перечислить, что не сделано

Далее в коде пишите комментарии, не слишком много, но чтобы было понятно, что в этом фрагменте (цикле, вводе и проч.) выполняется



Рекомендации

Желательно не использовать русский язык, в т.ч. в сообщениях, выводимых на консоль.

Желательно писать код, соответствующий стандартам C++ и компилирующийся под gcc (g++) — самые распространенные вещи:

- писать `int main()`, не писать `void main()`, хотя компилятор Microsoft это принимает;
- писать `#include <cstdlib>` для функций `rand()`, `system()` и т.д., хотя компилятор Microsoft включает это по умолчанию.



Консультации и защиты проектов

день	время	
Четверг, 13.07.2017	12-10 – 16-30	Консультации по проекту Защиты проектов
Пятница, 14.07.2017	12-10 – 16-30	Консультации по проекту Защиты проектов
Суббота, 15.07.2017 Только при крайней необходимости	12-10 – 15-00	Защиты проектов

Защита очная преподавателю || учебному ассистенту

На защите – вопросы по

- коду,
- алгоритмам,
- измерению времени,
- организации экспериментов,
- анализу и выводам по полученным результатам.



Отчет

В ЛМС загружаем один архив, содержащий

- Комментированный код (один файл *.crr или архив с проектом)
- Результаты эксперимента в электротаблице - таблицы и графики
- Отчет – постановка задачи, описание выполненной работы, анализ полученных результатов и выводы (текст)
- Имена файлов таблиц и текста - Pract<FIO>.*

Критерии проверки и оценивания проекта см. в таблице оценок





Возможные ошибки и недостатки

- нет таблиц и графиков;
- нет результатов и выводов ;
- неправильно работают или не реализованы алгоритмы сортировки
- Изменяется эталонный массив или повторно подаётся на вход алгоритму отсортированный на предыдущем шаге массив;
- Измерения для алгоритма отсутствуют или выполнены неправильно (не в (наносек. или тактах))
- Не создаётся CSV-файл с результатами измерений для дальнейшего создания и обработки таблиц;
- Используется дополнительная память там, где это не требуется;
- Наличие в функциях сортировок посторонних операций (копирование массивов, счетчики операций, вывод и т.п.);
- Нет проверки отрицательных/нулевых размеров массива;
- Нет проверки на буквы и прочий некорректный ввод;
- Нет рандомной генерации или не меняется seed (не используется srand);
- Неправильно генерируется массив какого-то типа
- Нет очистки памяти;
- Нет комментариев;



Возможные достоинства работы

- + использование указателей на функции в качестве параметров функций
- + полный и качественный анализ результатов
- + наличие всех графиков
- + отличные ответы на защите
- + все минусы предыдущего слайда можно обратить в плюсы, если сделать все верно
- + защита проекта досрочно (в четверг, 13.07.2017)

