

**Федеральное государственное автономное образовательное
учреждение высшего образования “Национальный
исследовательский университет
“Высшая школа экономики”
Московский институт электроники и математики им. А.Н. Тихонова
НИУ ВШЭ
Департамент компьютерной инженерии**

Курс: «Проектный семинар «Python в науке о данных»

**Техническое задание
по проекту “Political Coordinates”**

Группа:

БИБ225

Номер бригады:

Директор бригады:

Носов Иван Александрович

+7 (924) 301-27-51

ianosov@edu.hse.ru

Состав бригады:

Носов Иван Александрович,

Пантякова Дарья Евгеньевна,

Лифановский Дмитрий Валентинович

Руководитель:

Полякова Марина Васильевна

МОСКВА 2023

Содержание:

1 Введение

1.1 Наименование программы

1.2 Краткая характеристика области применения

2 Основания для разработки

3 Назначение разработки

3.1 Функциональное назначение

3.2 Эксплуатационное назначение

4 Требования к программе или программному изделию

4.1 Требования к функциональным характеристикам

4.1.1 Требования к составу выполняемых функций

4.1.2 Требования к организации входных и выходных данных

4.2 Требования к надежности

4.2.1 Требования к обеспечению надежного (устойчивого) функционирования программы

4.2.2 Время восстановления после отказа

4.2.3 Отказы из-за некорректных действий оператора

4.3 Условия эксплуатации

4.3.1 Климатические условия эксплуатации

4.3.2 Требования к видам обслуживания

4.3.3 Требования к численности и квалификации персонала

4.4 Требования к составу и параметрам технических средств

4.5 Требования к информационной и программной совместимости

4.5.1. Требования к исходным кодам и языкам программирования

4.5.2. Требования к программным средствам, используемым программой

4.6 Требование к маркировке и упаковке

4.7 Требования к транспортированию и хранению

4.8 Специальные требования

5 Требования к программной документации

6 Техничко-экономические показатели

7 Стадии и этапы разработки

7.1. Стадии разработки

7.2. Этапы разработки

7.3. Содержание работ по этапам

8 Порядок контроля и приемки

1 Введение

1.1 Наименование программы

Наименование программы – «Political Coordinates».

1.2 Краткая характеристика области применения

Специализированное информационно–аналитическое приложение «Political Coordinates» предназначено для сбора, обработки, хранения и анализа данных, полученных от респондентов или находящихся в базе данных. Анализ данных заключается в построении графических отчетов (в виде таблиц, диаграмм, гистограмм и пр.), анализа этих графических отчетов и их сравнения между собой.

Основная цель разрабатываемой программы - сделать работу социологов и других людей, работающих со статистическими данными, более быстрой и качественной.

2 Основания для разработки

Основанием для проведения разработки специализированного информационно–аналитического приложения является договоренность между преподавательницей Поляковой М. В. (далее Заказчик) и группой студентов, включающей Носова И. А., Пантякову Д.Е. и Лифановского Д. В. (далее Исполнитель).

Согласно договоренности, Исполнитель обязан разработать систему «Political Coordinates» на курсе «Проектный семинар «Python в науке о данных» и продемонстрировать ее работоспособность и достигнутые результаты. Окончательная версия информационно-аналитического приложения, «Руководства пользователя» и «Руководства разработчика» публикуется в Telegram-группе директором бригады не позднее, чем за 10 дней до начала сессии.

3 Назначение разработки

3.1 Функциональное назначение

Программа предоставляет возможность пройти тест на определение политических координат, построить графические отчеты на основе полученных после теста параметров и на основе данных, полученных от всех респондентов. Также существует возможность провести сравнительный анализ двух разных категорий по разным ответам и получить научную справку по наименованиям, понятиям и терминам, являющимися профессиональными в социально-гуманитарных науках.

Анализ данных осуществляется посредством создания и демонстрации пользователю следующих отчетов:

- Простой текстовый отчет, для создания которого необходимо применить проекцию и сокращение таблицы путем удаления лишних строк и столбцов. Данную операцию нужно выполнить при помощи библиотеки Pandas, которая содержит инструменты объединения таблиц, что позволит сократить таблицу и упростить ее представление в отчете;
- Текстовый статистический отчет по любому набору атрибутов (набор основных описательных статистик). Отчет включает в себя:
 - для качественных переменных включает таблицу частот. Первый столбец таблицы содержит значения переменной, второй столбец содержит частоты, третий столбец содержит процент количества указанных объектов от их общего числа;
 - для количественных переменных включает основные статистики. Отчет должен быть оформлен в виде таблицы, в которой первый столбец содержит список переменных, далее в каждом столбце значение статистики, которое подсчитывается с помощью методов класса `pandas.DataFrame`;
- Текстовый отчет «сводная таблица» для любой пары качественных параметров с возможностью выбора метода агрегации. Построение отчета возможно при использовании функции `pandas.pivot_table()`;
- Графический отчет «кластеризованная столбчатая диаграмма» для пары «качественный атрибут — качественный атрибут». При создании отчета следует использовать `matplotlib.pyplot.bar()`;
- Графический отчет «категоризованная гистограмма» для пары «количественный атрибут — качественный атрибут», который строится при помощи `matplotlib.pyplot.hist()`;
- Графический отчет «категоризованная диаграмма Бокса-Вискера» для пары «количественный атрибут — качественный атрибут». Для его создания необходимо воспользоваться `matplotlib.pyplot.boxplot()`;
- Графический отчет «категоризованная диаграмма рассеивания» для двух количественных атрибутов и одного качественного атрибута, построение которого происходит посредством использования `matplotlib.pyplot.scatter()`.

Для всех текстовых отчетов следует предусмотреть опцию вывода в файл. Все графические отчеты могут быть экспортированы в графические файлы.

3.2 Эксплуатационное назначение

Программа предназначена для эксплуатации социологами и другими людьми в профессиональной среде для ускорения, совершенствования и повышения точности анализа большого количества данных, полученных от респондентов в ходе социологических опросов.

Конечным пользователем программы может являться и человек, не работающий в профессиональной среде. В таком случае приложение может

использоваться в развлекательных целях и с целью популяризации активной гражданской позиции среди населения.

4 Требования к программе или программному изделию

4.1 Требования к функциональным характеристикам

4.1.1 Требования к составу выполняемых функций

Программа должна предоставлять возможность прохождения теста на определение политических координат, получения определенных параметров в качестве результата и занесение новых полученных статистических данных в базу данных.

Также должна осуществляться работа с базой данных для добавления новых записей и столбцов и их удаления, загрузки и сохранения базы данных. Добавление, удаление и модификация объектов в базе данных должны иметь возможность осуществляться ручным способом. База данных содержит 8 основных полей, а именно:

1. id - индивидуальный номер каждой записи;
2. gender - содержит информацию о гендере респондента;
3. field - научное направление, которое глобально характеризует ОП респондента;
4. university - учебное заведение респондента;
5. course - курс обучения респондента;
6. x - содержит координату по оси X, которая была ранее вычислена определенным способом и характеризует политические предпочтения респондента;
7. y - аналогично столбцу "x", только по оси Y;
8. z - аналогично столбцу "x", только по оси Z.

Параллельно с этим должна существовать возможность построения всех типов текстовых и графических отчетов, на основании всех записей, находящихся в базе данных, или на основании групп данных, выбранных определенным способом. А именно:

- Простой текстовый отчет;
- Текстовый статистический отчет;
- Текстовый отчет «сводная таблица»;
- Графический отчет «кластеризованная столбчатая диаграмма»;
- Графический отчет «категоризированная гистограмма»;
- Графический отчет «категоризированная диаграмма Бокса-Вискера»;
- Графический отчет «категоризированная диаграмма рассеивания».

Подробно все типы отчетов описаны выше в разделе "3.1 Функциональное назначение".

Вдобавок может осуществляться получение научной справки по определенным малоизвестным, профессиональным терминам.

Все описанные выше пункты, этапы, разделы и проект в целом, осуществимы только при разработке следующих модулей:

- модуль работы с базой данных: добавление-удаление записей, столбцов, загрузка и сохранение бд;
- модуль создания отчетов: сохранение бд с удаленными записями/столбцами (файл таблицы), создания отчетов, описанных ранее;
- модуль разработки графического интерфейса (gui): построение графических отчетов и прохождение теста.

4.1.2 Требования к организации входных и выходных данных

Статистические данные о респондентах, уже прошедших опрос, хранятся в базе данных. Следуя критериям, эта база данных должна быть в двоичном формате. Например, база данных должна храниться в файле с расширением .csv.

Все выходные данные осуществляются при помощи графического интерфейса в виде графических и текстовых отчетов, описанных ранее.

4.2 Требования к надежности

4.2.1 Требования к обеспечению надежного (устойчивого) функционирования программы

У пользователя, работающего с программой, должны быть установлены все библиотеки, согласно конфигурационному файлу, для корректной реализации работы программы. В дополнение к этому приложение должно иметь постоянный, бесперебойный доступ к базе данных. Одним из основных требований к надежному функционированию является организация бесперебойного питания устройства, на котором происходит запуск программы, а также использование лицензионного ПО на нем.

4.2.2 Время восстановления после отказа

Время на восстановление после отказа, вызванного неисправностью технических средств, фатальным сбоем операционной системы, не должно превышать времени, необходимого на устранение неисправностей технических средств и переустановки программных средств.

4.2.3 Отказы из-за некорректных действий оператора

Отказы программы возможны вследствие некорректных действий оператора при взаимодействии с операционной системой и программным обеспечением.

4.3 Условия эксплуатации

4.3.1 Климатические условия эксплуатации

Климатические условия эксплуатации, при которых должны обеспечиваться заданные характеристики, должны удовлетворять требованиям, предъявляемым к техническим средствам согласно условиям их эксплуатации.

4.3.2 Требования к видам обслуживания

Вся информация находится в разделе "Требования к обеспечению надежного (устойчивого) функционирования программы".

4.3.3 Требования к численности и квалификации персонала

Для управления системой достаточно 2 человек: местный системный администратор и пользователь, если система используется в профессиональном сегменте.

В перечень задач, выполняемых системным администратором, должны входить:

- задача поддержания работоспособности технических средств;
- задачи установки и поддержания работоспособности системных программных средств - операционной системы;
- задача установки/запуска программы.

К квалификации пользователя специальные требования не предъявляются.

Количество человек для управления системой может быть сокращено до 1, если пользователь обладает навыками, описанными выше.

4.4 Требования к составу и параметрам технических средств

Вся реализация программы осуществляется непосредственно на устройстве пользователя. В состав технических средств должен входить персональный компьютер.

Его минимальные характеристики:

- процессор x86 с тактовой частотой, не менее 1 ГГц;
- оперативная память объемом, не менее 1 Гб;
- монитор, мышь, клавиатура;
- ОС MS Windows 10.

4.5 Требования к информационной и программной совместимости

4.5.1. Требования к исходным кодам и языкам программирования

Исходные коды программы должны быть написаны на языке программирования Python3 при помощи дистрибутива Anaconda.

4.5.2. Требования к программным средствам, используемым программой

Системные программные средства, используемые программой, должны быть представлены лицензионной версией операционной системы Windows 10.

Основой для системы должна стать база данных, в которой будет храниться вся информация.

В системе должны быть установлены следующие библиотеки для корректной работы программы:

- numpy 1.24.3
- pandas 2.0.1
- python-dateutil 2.8.2
- pytz 2023.3
- six 1.16.0
- tzdata 2023.3

Этот список указан в файле requirements.txt.

Разрабатываемой приложение имеет определенную структуру и, следуя ей, размещается в стандартной структуре каталогов:

- Work - основной каталог;
 - Data - содержит базу данных;
 - Library - содержит библиотеку стандартных функций;
 - Notes - содержит документацию;
 - Script - содержит специализированный модуль и файл с определением параметров настройки приложения.

4.6 Требование к маркировке и упаковке

Программа предоставляется строго в виде архива каталогов «work», содержащих все компоненты проекта. Программное изделие должно иметь маркировку с обозначением архива <Группа (три символа)> _ <Бригада (два символа)> _ <Версия (три символа)>.zip. В остальных случаях требования к маркировке и упаковке не предъявляются.

4.7 Требования к транспортированию и хранению

Доступ к архиву осуществляется через Telegram-группу. В остальном, специальных требований не предъявляется.

4.8 Специальные требования

Программа должна обеспечивать взаимодействие с пользователем посредством графического пользовательского интерфейса, разработанного согласно рекомендациям, описанным в документе “Описание самостоятельной работы по курсу — “Проектный семинар “Python в науке о данных”.

5 Требования к программной документации

Документы к программе должны быть выполнены в соответствии с ГОСТ 19 и ГОСТами к каждому виду документа.

Предварительный состав программной документации:

- Техническое задание (включает описание применения);
- Пояснительная записка;
- Руководство оператора (руководство пользователя);
- Руководство программиста (руководство разработчика);
- Программа и методика испытаний;
- Текст программы.

6 Техничко-экономические показатели

Использование разрабатываемого приложения сократит время,затрачиваемое на обработку и анализ данных, полученных от респондентов в ходе социологического опроса. Вдобавок программа сократит время на ознакомление с терминами предметной области. В совокупности это позволит повысить эффективность работы научных сотрудников.

7 Стадии и этапы разработки

Дата	Контрольная точка
01.02.2023	Согласование темы проекта
15.03.2023	Представление первых графиков
10.05.2023	Демонстрация статистических отчетов
25.05.2023	Представление финального технического задания и графического интерфейса
11.06.2023	Отправка окончательной версии всего проекта
21.06.2023	Защита проекта

В ходе создания и разработки проекта было примерно подсчитано общее затраченное время. В итоге на всю работу над проектом суммарно было затрачено около 60 часов.

7.1. Стадии разработки

Разработка должна быть проведена в три стадии:

1. техническое задание;
2. технический и рабочий проекты;
3. защита перед комиссией.

7.2. Этапы разработки

На стадии «Техническое задание» должен быть выполнен этап разработки, согласования и утверждения данного технического задания.

На стадии «Технический и рабочий проект» должны быть выполнены перечисленные ниже этапы работ:

- разработка программы;
- разработка программной документации;
- испытания программы.

7.3. Содержание работ по этапам

На этапе разработки технического задания должны быть выполнены перечисленные ниже работы:

- постановка задачи;
- определение и уточнение требований к техническим средствам;
- определение требований к программе;
- определение стадий, этапов и сроков разработки программы и документации на нее;
- согласование и утверждение технического задания.

На этапе разработки программы должна быть выполнена работа по программированию (кодированию) и отладке программы.

На этапе разработки программной документации должна быть выполнена разработка программных документов в соответствии с требованиями ГОСТ 19.

На этапе испытаний программы должны быть выполнены перечисленные ниже виды работ:

- разработка, согласование и утверждение порядка и методов проведения испытаний;
- проведение приемо-сдаточных испытаний;
- корректировка программы и программной документации на основании результатов испытаний.

На этапе защиты перед комиссией должна быть выполнена работа по подготовке и передаче программы и программной документации заказчику и членам комиссии.

8 Порядок контроля и приемки

Приемосдаточные испытания программы должны проводиться согласно разработанной исполнителем и согласованной заказчиком «Программы и методики испытаний».

Ход проведения приемо-сдаточных испытаний заказчик и исполнитель документируют в протоколе испытаний.

На основании протокола испытаний исполнитель совместно с заказчиком подписывают акт приемки-сдачи программы в эксплуатацию.

Вдобавок контроль и приемка разработки осуществляются в соответствии с документом «Описание самостоятельной работы по курсу — «Проектный семинар «Python в науке о данных»».