# Задание 5. Разработка собственного проекта

## Задание

В ходе этого задания вам необходимо описать проблемную ситуацию, которую можно решить с помощью анализа текстов, и разработать проект решения данной проблемы.

Задание разделено на несколько этапов, в конце каждого этапа группа должна предоставить модератору промежуточные результаты. Все этапы перечислены ниже.

## Этапы работы

### 1.Описание проблемы, постановка задачи

#### Задание

В ходе обсуждения выберите и опишите проблему, которую вы будете решать с помощью анализа текстов.

На основе описанной проблемы сформулируйте постановку задачи (поиск похожего текста, кластеризация, или при желании классификация или регрессия).

Опишите сценарий (сценарии) решения проблемы с помощью вашего продукта: что пользователь вашего продукта должен сделать, чтобы решить описанную проблему. При описании сценариев учитывайте, что у пользователей вашего продукта могут быть разные роли.

Опишите данные, которые планируется использовать в ходе разработки вашего продукта, а также источник этих данных, специальные требования (метаданные), предполагаемый достаточный объем.

#### Ожидаемые результаты

* Описание проблемы
* Постановка задачи
* Описание сценариев работы с вашей разработкой (с учетом ролей пользователей)
* Описание данных и требований

#### Контрольная точка

**17 июля.** Файл с описанием по списку ожидаемых результатов (проблема, задачи, сценарии, описание данных и требований).

### 2.Уточнение проблемы и постановки задачи, сбор данных

#### Задание

Выполните предварительный сбор корпуса текстов, чтобы оценить доступность данных, их качество, объем. При необходимости доработайте постановку задачи.

Исходя из представления о данных сформулируйте предполагаемые этапы решения задачи - этапы сбора, предобработки и построения модели. Опишите особенности ваших данных.

#### Контрольные точки

**20 июля.** 2 файла: Описание требований к данным, Собранный корпус текстов.

**21 июля.** Защита по итогам первого этапа: проблема, актуальность, постановка задачи, сценарии использования, планируемые этапы разработки решения, описание и примеры данных.

### 3.Описание решения

#### Задание

Подготовить описание решения поставленной задачи. Решение включает в себя:

* перечень инструментов: библиотеки, готовые программные продукты (например, парсеры или выгрузка);
* поэтапное описание того, что нужно сделать, чтобы решить задачу;
* описание эксперимента и результаты целью которого является проверка работоспособности вашего решения, либо проверка гипотезы относительно работоспособности модели в определенных условиях; может быть описание поэтапной реализации решения с выводами по каждому этапу и обоснованием принятия промежуточных решений;
* работающая модель;
* прототип (У слушателей должна быть возможность ввода своего текста и проверки результатов. Лучше также заготовить несколько примеров, чтобы можно было на демонстрации прототипа выбрать один из примеров и посмотреть на результат).

#### Ожидаемые результаты

* Перечень инструментов
* Этапы разработки решения
* План эксперимента
* Результаты эксперимента
* Прототип работающей модели

#### Контрольная точка

**22 июля. Важно:** 21 июля и 22 июля необходимо подводить промежуточные итоги вашей работы и отправлять результат модератору для загрузки в систему.

### 4.Подготовка презентации

#### Задание

Подготовить презентацию проекта, включающую в себя проработанные этапы:

* описание проблемы
* постановка задачи
* сценарии применения
* описание данных
  + требования и общее описание данных
  + примеры данных
* предлагаемое решение
  + описание решения, особенности
  + описание экспериментов (с результатами)
  + результаты работы прототипа

#### Контрольная точка

**24 июля.** Защита проекта.