Задание 1 - Необходимо выделить основных действующих лиц и организации, а также определить суть события, описываемого в новости.

Введение: Для решения поставленной задачи необходимо:

1. Распарсить данные из xlsx файла.
2. Разделить текст на предложения.
3. Определить действующих лиц в предложении.
4. Построить граф показывающий отношение, действия между лицами.

Основная часть:

1. Алгоритм выделения активных субъектов (людей и организаций) в тексте:
   1. Выделим из текста предложения по знакам пунктуации.
   2. Определим части речи для каждого из слов.
   3. По частям речи определяем смысловые куски
   4. Визуализируем отношения при помощи графа
2. Суть события выявляется путем сопоставления графа с узлом (действием глагол) и примыкающих к нему субъектов.
3. Выбор инструментов (языка, библиотек, моделей) для реализации алгоритмов:

Основой для программы будет язык Python 3.8. Это язык, для которого есть много модулей для обработки информации, в том числе и NLP. Для парсинга файла была выбрана удобная библиотека pandas. Обработка текста выполняется с помощью модуля spacy. Я выбрал этот модуль, так как он достаточно прост в использовании и имеет русскоязычную модель. Построение графа осуществляется при помощи networkx. Отрисовкой занимается классический модуль matplotlib.

Заключение: В результате выполнения задания была получена программа, отображающая взаимосвязи между объектами через действия с ними.

Ссылка на проект: <https://github.com/ndmarinin/Task1-YP>   
Запуск: установить библиотеки из requirements.txt и запустить main.py

После идет обработка файла news\_dataset.xlsx размещенный в директории проекта. В консоли нужно ввести номер новости для ее обработки. После обработки можно выбрать следующий номер.