Необходимо создать нейросеть, которая станет решением для неточного поиска (когда поиск идет не по полному соответствию фразы, а по ее смыслу). Поиск осуществляем по Википедии, по базе данных событий, по запросу в поисковую строку Google и по новостным сайтам.

В процессе обучения нейронки необходимо расширить ее возможности до прогнозирования событий в тот или иной день будущего, на основе расчета повторяемости тех или иных событий в разные дни прошлого и расчета «предрасположенности» (дополнительных факторов) в настоящем, которая теоретически увеличивает вероятность того или иного события в конкретный день. Например, при военных действиях возрастает риск терактов, при длительном застаивании тектонических плит повышается вероятность землетрясений и т.д. Также необходимо учитывать ожидания специалистов в конкретной области. Например, ученые ожидают землетрясение в том или ином районе - это повышает вероятность. Также нейросеть может учесть и другие факторы, не обозначенные выше.

Заказчик вводит определенный ключевой запрос и интервал времени, а алгоритм выдает все события, подходящие под ключевой запрос, за этот период, рассчитывая количество дней между событиями. Далее нейросеть должна рассчитать вероятность события, подходящего под критерии ключевого запроса, в определенный день. Третий вид запросов – когда Заказчик вводит определенную дату из будущего без ключа, а нейросеть расписывает вероятность тех или иных событий в эту дату.

1. Для выполнения задачи необходимо спарсить раздел Википедии Current Events с января 2002: <https://en.wikipedia.org/wiki/Category:Current_events_archives>

Необходимо спарсить наименование события и дату в каждой из вкладок. Сайт на английском языке, так что нужен автопереводчик.

Списки впоследствии будут обновляться - нужны данные в актуальном времени.

1. Для выполнения задачи необходимо применить код по другим трем разделам Википедии. А также по сайтам с базой данных по терактам, природным и технологическим событиям, а также по падающим метеоритам (данные с за последние 50 лет). Коды будут высланы Исполнителю дополнительно.
2. Для выполнения задачи необходимо учесть в выборке следующие новостные сайты:

<https://www.reuters.com/>

<https://www.theguardian.com/international>

<https://news.sky.com/>

<https://edition.cnn.com/>

<https://www.nytimes.com/>

<https://www.washingtonpost.com/>

<https://www.foxnews.com/>

<https://news.yahoo.com/>

<https://nypost.com/>

<https://www.interfax.ru/>

<https://www.rbc.ru/>

Большинство сайтов на английском языке, так что нужен автопереводчик.

1. Необходимо разработать скрипт для обращения нейросетки к агрегатору Google.

Формат вывода данных – txt

Каждый запрос должен формировать ОДИН файл txt, в название которого включен ключ/ключевой запрос.

Дедлайн проекта не установлен.

В процессе работы над проектом могут возникнуть «подводные камни», Заказчик обязуется всегда быть на связи и отвечать на возникающие вопросы Исполнителя. Кроме того, Заказчиком в процессе могут быть сформулированы небольшие дополнения/уточнения, которые не будут входить в противоречие с изначально сформулированной задачей. Эти дополнения/уточнения не должны отразиться на изначально оговоренной стоимости проекта, которая составит **6000 рублей**.

**ПРИМЕРЫ**

Запрос 1: Пожар во Франции январь 1990 – декабрь 2023 .

1. *Пожар в туннеле под Ла-Маншем 18 ноября 1996.*

*Интервал - … дней*

1. *Пожар в Монбланском туннеле 24 декабря 1999.*

*Интервал - … дней*

1. *Пожар в Нотр-дам-де-Пари 15 апреля 2019*

*И т.д.*

Запрос 2: Пожар во Франции 15 апреля 2024

*Вероятность: 0,32%.*

*Всего проанализировано n событий.*

*Всего проанализировано n факторов.*

Запрос 3: 18 ноября 2024

*Землетрясение: 0,5% (n событий, n факторов)*

*Извержение вулкана: 1,35% (n событий, n факторов)*

*Теракт: 2,15% (n событий, n факторов)*

*Пожар: 1,55% (n событий, n факторов)*

*И т.д.*