**Перечень вопросов**

**В билете должно быть 2 теоретических и 1 практический вопрос.**

**Теоретические вопросы:**

**Вопрос 1**. Основные этапы предварительной обработки текста. Удаление нерелевантных символов, числовая кодировка текстовой информации, снижение уникальности, токенизация.

**Вопрос 2.** Токенизация текста. 6 основных характеристик Tokenizer в tensorflow.

**Вопрос 3.** Токенизация теста. Преобразование текста в вектор. text\_to\_matrix(), texts\_to\_sequences() отличия, условия применения. Примеры преобразования текста. Переход от векторной формы к словам.

**Вопрос 4.** Сеть прямого распространения. Принцип работы, основные понятия. Архитектура сети прямого распространения. Dense слой.

**Вопрос 5.** Полносвязный слой нейронной сети. Слой Dropout. Функция активации.

**Вопрос 6.** 4 основных оптимизатора keras. Принципы работы. Этап обучение нейронной сети на котором осуществляется выбор оптимизатора.

**Вопрос 7.** Эмбеддинг при анализе текстов. Цель. История развития. Допустимые операции при токенизации при использовании эмбеддинга. Гиперпараметры слоя Embedding Keras.

**Вопрос 8.** Требования к архитектуре нейронных сетей с Embedding слоем. Принцип работы SpatialDropout1D, Flatten.

**Вопрос 9.** Недостатки Dense слоя. Рекуррентный нейрон, шаги обучения. Основная идея рекуррентной нейронной сети. Гиперпараметры простого нейронного слова в Keras. Задачи рекуррентных нейронных сетей.

**Вопрос 10.** Рекуррентные нейронные сети. Проблема исчезающего градиента. LSTM слой, структура, основные блоки.

**Вопрос 11.** LSTM-сети. Задачи. Работа слоя потери. Параметры W, h, x.

**Вопрос 12.** LSTM-сети. Задачи. Слой сохранения. Вектор контекста, параметр С.

**Вопрос 13.** LSTM-сети. Задачи. Слой обновления. Вектор контекста. Гиперпараметры: units, activation, return\_sequences, recurrent\_activation.

**Вопросы 14.** Двунаправленные рекуррентные сети. Bidirectional.

**Вопрос 15.** Рекуррентная нейронная сеть. GRU. Отличия от LSTM.

**Вопрос 16.** Слои нейронной сети. Conv1D. Окно свертки, ядро свертки.

**Вопрос 17.** Задача NER. Типовые сущности. Метки начала и окончания сущностей. Метод перехода к векторной форме при решении задачи NER при токенизации. Варьируемые параметры при обучении нейронной сети для решения задачи NER

**Вопрос 18.** Задача NER. Постановка задачи. Вид признаков и целевого массива. Метод кодирования целевого признака. Пример архитектуры нейронной сети при решении задачи NER.

**Вопрос 19.** Лемматизация текста. Инструменты, основные библиотеки.

**Вопрос 20.** Работа с многомерными данными. Процесс перехода от текста к многомерным данным. Свертка. Ядро свертки, движение ядра по исходной матрице. Max Pooling, Avg Pooling. Фильтры при свертке n мерных данных. Особенности Conv1D

**Вопрос 21.** Слои нейронной сети. Conv1D, SpatialDropout1D. Гиперпараметры слоев, **inputs, outouts**. Результаты преобразования.

**Вопрос 22.** Модели Seq2Seq. Задачи. Общая структура. Encoder и Decoder. Вектор контекста

**Вопрос 23.** Модели Seq2Seq. Encoder. Пример архитектуры нейронной сети. Зависимость параметров слоя Embedding и LSTM. Пример данных, возвращаемых Encoder. Характеристики данных, подаваемых на вход. Объединение Encoder и Decoder, обучение Seq2Seq.

**Вопрос 24.** Модели Seq2Seq. Decoder. Пример архитектуры нейронной сети. Имя слоев нейронной сети. Характеристики данных, подаваемых на вход. Decoder output. Объединение Encoder и Decoder, обучение Seq2Seq.

**Вопрос 25.** Модели Seq2Seq. Общая архитектура нейронных сетей. Обучение модели Seq2Seq. Модель формирования ответа.

**Вопрос 26.**  Модели Seq2Seq. Общие проблемы классической Seq2Seq. Seq2Seq, общий принцип работы. Отличие от простой Seq2Seq. Вектор внимания, вектор контекста. Размерность вектора контекста.

**Вопрос 27.** Модели Seq2Seq с вниманием. Encoder. Пример архитектуры нейронной сети. Гиперпараметры рекуррентной нейронной сети, обеспечивающий работу сети с вниманием. Отличие результатов работы рекуррентного слоя простой Seq2Seq от Seq2Seq с вниманием.

**Вопрос 28.** Модели Seq2Seq с вниманием. Decoder. Пример архитектуры нейронной сети. Размерность данных, получаемых от Encoder и Decoder.

**Вопрос 29.** Трансформеры при анализе тестовых данных. Общий принцип работы. Общая структура Encoder и Decoder.

**Вопрос 29.** Трансформеры при анализе тестовых данных. Вычисление внутреннего внимания. Три основных параметра.

**Вопрос 30.** Получение данных из обученных токинайзеров. Библиотеки для лемматизации. Получение нормальной формы слова.

**Вопрос 31.** Типы нейронных сетей. Сеть прямого распространения (Sequential) и сети с несколькими входами и выходами (для Seq2Seq). Примеры задание сеть на основе Sequential и на основе Model.

**Вопрос 32.** Текстовые модели как основа поисковых систем. Графы знаний. Word2vec, skip-gram, GloVe – Global

**Вопрос 32.** Текстовые модели как основа поисковых систем. Переход от сетевого графа к графу знаний. Синтаксический анализ. Анализ сущностей.

**Вопрос 33.** Тематическое моделирование. LDA.

**Вопрос 34.** Архитектуры нейронных сетей, использующихся для анализа текстов. RNN, прямая и обратная модель. Встраиваемая языковая модель. ELMo.

**Вопрос 35.** Суммаризация текстов. Виды. TextRank. Метод Луна. LSA.

**Вопрос 36.** LLM. История возникновения и развития. Использование предобученных. Промт. Тюнинг моделей.

**Практические вопросы:**

1. Вам необходимо импортировать CSV-файл с названием “Spam-сообщения”.

После этого, необходимо произвести предварительную обработку текстового массива, после чего вы должны произвести обучение модели для задачи классификации.

1. Вам необходимо импортировать CSV-файл с названием “Тексты писателей”.

После этого, необходимо произвести предварительную обработку текстового массива, после чего вы должны произвести операции по анализу тональности текста и построить граф только с позитивным окрасом. Назовите 3 вершины графа с самыми крепкими связями

1. Вам необходимо импортировать CSV-файл с названием “Тексты писателей”.

После этого, необходимо произвести предварительную обработку текстового массива, после чего вы должны произвести операции по анализу тональности текста и построить граф только с негативным окрасом. Назовите 3 вершины графа с самыми слабыми связями

1. Вам необходимо импортировать CSV-файл с названием “Тексты писателей”.

После этого, необходимо произвести предварительную обработку текстового массива, после чего вы должны произвести операции по анализу тональности текста и построить граф. Назовите 3 вершины графа и назвать количественный показатель их плотности связи

Ссылки:

<https://habr.com/ru/articles/482126/>

<https://habr.com/ru/articles/486358/>

https://github.com/natasha/slovnet

https://habr.com/ru/company/wunderfund/blog/331310/