

Выявление оптимальной длины запроса к модели по собранным данным из опроса.

Цель опроса:

Собрать датасет возможных вопросов от студентов разных курсов по информации размещённой на сайте se.moevm.info.

Предметная область.

В качестве предметов для опроса были выбраны дисциплины первого курса, а именно “Программирование” и “Информатика”, поскольку у первокурсников как правило появляется больше вопросов.

Способ сбора информации.

Опрос был составлен в виде гугл-формы и распространён по электронной почте студентам первого курса ЛЭТИ.

Ссылка на гугл форму:

<https://docs.google.com/forms/d/e/1FAIpQLSfJjsfsdbpF9t4ZLxlcL0sUTPGL48-J0R4ZyQYy9Spkuob2tA/viewform>

Результаты опроса (14.05.24)

Полученные данные от студентов представлены в Таблице 1.

Таблица 1. Результаты опроса.

1 курс. Программирование.	Длина запроса	1 курс. Информатика.	Длина запроса
Какие темы лекций?	18 симв. 3 слова	Чем можно пользоваться на защите лабораторных работ?	52 симв. 7 слов
Как узнать вариант курсовой работы?	35 симв. 5 слов	Как получить 5 (отлично) по предмету Информационные технологии?	63 симв. 8 слов
Как можно связаться с преподавателем по программированию?	57 симв. 7 слов	Где можно посмотреть презентации с лекций?	42 симв. 6 слов
Как защитить курсовую работу, если ты почти ничего не знаешь?	61 симв. 10 слов	Когда лекция по информатике?	28 симв. 4 слов
Какая почта у Иванова Дмитрия?	30 симв. 5 слов	Какие темы для весеннего семестра?	34 симв. 5 слов
Какие требования к курсовой работе?	35 симв. 5 слов	Темы первой контрольной работы	30 симв. 4 слов
Какая документация есть по этому предмету?	42 симв. 6 слов	По каким темам будут лабораторные работы?	41 симв. 6 слов
Будет ли контрольная или только лабораторные?	45 симв. 6 слов	Какие есть простые задачи для повышения баллов?	47 симв. 7 слов
Как оформить письмо преподавателю?	34 симв. 4 слова	В каком стиле оформить "Листинг"?	33 симв. 5 слов
Что необходимо сделать, чтобы защитить курсовую	66 симв. 9 слов	Опиши правила коммуникации по электронной почте.	48 симв. 6 слов

и получить оценку?			
--------------------	--	--	--

Анализ полученных результатов.

Согласно данным о средней скорости речи на русском языке [1], человек произносит от 80 до 120 слов в минуту. Взяв нижнюю границу получим примерно 1.33 слова в секунду. Максимальное количество слов в полученных вопросах - 10. Следовательно оптимальное ограничение будет $10 / 1.33 = 7.52$. Округлим в большую сторону и получим 8 секунд чистой речи.

Также возьмём во внимание, что человек не сразу начнёт говорить и может задержаться в конце, пока нажимает на кнопку отправки или соответствующей клавиши на клавиатуре. Поэтому мы считаем, что стоит добавить примерно 2 секунды для обоих действий. Итоговое оптимальное ограничение времени аудиосообщения - 12 секунд.

Длина текста полученных вопросов составляет максимум 66 символов с учётом пробелов, в среднем - 42 символа. Точно определить оптимальную длину возможного вопроса по любому из контекстов не представляется возможным. Поэтому предлагаем ограничение несколько превышающее наибольший результат - 100 символов.

Выводы.

Таким образом, был собран и проанализирован датасет возможных вопросов по информации, размещённой на сайте se.moevm.info. Были оценены оптимальные ограничения для длины запроса модели ассистента, 12 секунд для аудио-сообщения и 100 символов для текстового сообщения.

Источники.

1. Большая Российская энциклопедия 2004 - 2017 [Электронный ресурс]. URL: <https://old.bigenc.ru/linguistics/text/4186876>