Общая информация о программе

# Суммаризатор

## Для чего нужна суммаризация текста?

Суммаризатор текста — это инструмент, который использует искусственный интеллект и сложные алгоритмы для преобразования текста из его длинной и подробной версии в краткую и понятную. Инструмент суммирования переносит все ключевые моменты текста в сжатую версию. Он создает краткие изложения и выделяет главные тезисы.

## Кому он будет полезен?

Суммаризатор будет полезен для создания аннотаций научных докладов, статей, поможет сжать большое количество текста, выделяя ключевые тезисы и сохраняя смысл для удобства и упрощения работы студентам, журналистам, блогерам, копирайтерам и другим.

## Как это работает?

Пользователь онлайн отправляет в суммаризатор оригинальный исходный текст со стандартными параметрами генерации, которые при необходимости можно изменить. Суммаризатор генерирует варианты на сокращение, выбирает из них лучший и возвращает финальный сжатый текст с основными тезисами, в том числе перефразированными.

Нейросети были обучены на базе генеративной модели ruT5 на большом количестве материалов разной длины, стилистики и жанров. Их можно использовать для адаптации любого текстового контента, который нужно сократить или изложить другими словами, сохранив значение.

# Как использовать сервис?

1. Перейти на сайт.

Изображение выглядит как текст, снимок экрана, дизайн

Автоматически созданное описание

1. Выбрать суммаризатор и нажать на кнопку «Перейти».

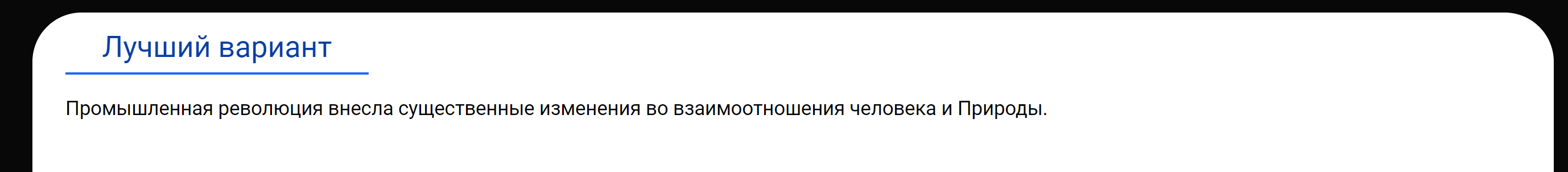
Изображение выглядит как текст, снимок экрана, мультимедиа, программное обеспечение

Автоматически созданное описание

1. Ввести свой текст в контекстное поле, выбрать режим пользования (при необходимости изменить настройки) и нажать «Сжать текст».

Изображение выглядит как текст, снимок экрана, программное обеспечение, Операционная система

Автоматически созданное описание



# Рерайтер

## Для чего нужен рерайтинг текста?

Рерайтинг текста- обработка исходного текста, путем изменения слов для создания нового текста с прежним смыслом.

## Кому он будет полезен?

Рерайтинг позволяет писать новости, продвигать сайты через SEO, заполнять информацию для товаров и услуг, размещать контент у партнеров.

Рерайтер будет полезен журналистам, репортерам, владельцам веб-ресурсов, менеджерам.

## Как это работает?

Базовое использование модели подразумевает работу с уже обученной моделью в режиме инференса. Модель умеет переписывать переданный текст. На вход подается оригинальный текст с настраиваемыми параметрами генерации, рерайтер генерирует варианты, выбирает из них лучший и возвращает финальный рерайт. Пользователь может выбрать кол-во генерируемых параметров, а также стратегию выбора лучшего кандидата.

Рерайтер был обучен с помощью [ruT5-large](https://huggingface.co/sberbank-ai/ruT5-large/tree/main) на текстах разной длины (как на уровне предложений, так и текстовых параграфов), из источников: новостных, разговорных, отзывов, художественной литературы; также были добавлены данные из открытых источников (отфильтрованные [tapaco](https://huggingface.co/datasets/tapaco), ParaphraserPlus).

# Как использовать сервис?

1. Перейти на сайт.

Изображение выглядит как текст, снимок экрана, программное обеспечение, электроника

Автоматически созданное описание

1. Выбрать рерайтер и нажать на кнопку «Перейти».

Изображение выглядит как текст, снимок экрана, мультимедиа, программное обеспечение

Автоматически созданное описание

1. Ввести свой текст в контекстное поле, выбрать режим пользования (при необходимости изменить настройки) и нажать «Переписать текст».

Изображение выглядит как текст, снимок экрана, программное обеспечение, Операционная система

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

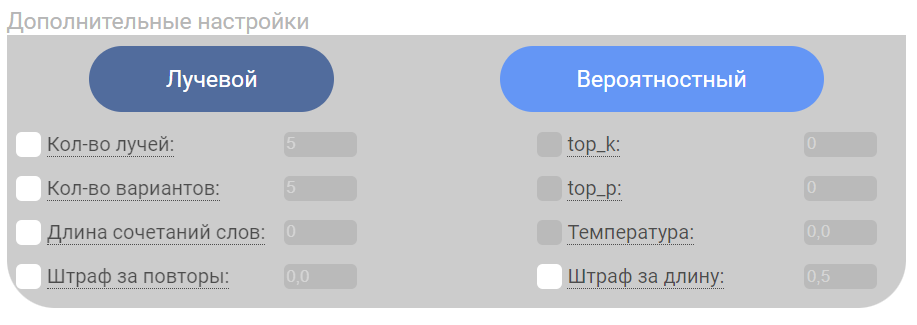
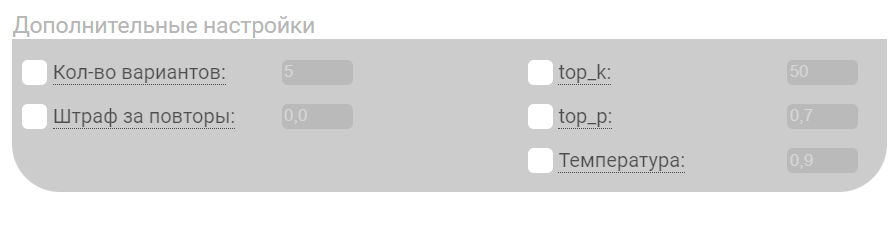
Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

# За что отвечают методы?

* **Лучевой** (Beamsearch). Эта стратегия оценивает несколько гипотез на каждом временном шаге и в конечном итоге выбирает гипотезу, имеющую наибольшую вероятность для всей последовательности. Преимуществом такой стратегии является выявление высоковероятных последовательностей, которые начинаются с начальных лексем с меньшей вероятностью и были бы проигнорированы при жадном поиске.
* **Вероятностный** (Sampling). Стратегия, при которой следующее слово выбирается случайно, исходя из вероятностей всех слов занять его место. Текст при таком методе является более интересным и оригинальным, однако, чаще возникают лексические ошибки.

# За что отвечают настройки?

* **Кол-во лучей** (num\_beams). Этот параметр показывает количество гипотез на каждом временном шаге. Это целое число больше 1.
* **Кол-во вариантов** (num\_return\_sequences). Количество последовательностей, которые вернёт ИИ в результате. Из этих последовательностей ИИ выбирает лучшую. Параметр является целым числом. Для метода генерации Beamsearch значение не должно превышать значение параметра «Кол-во лучей».
* **Длина сочетаний слов** (no\_repeat\_ngram\_size). Этот параметр показывает количество слов в сочетаниях слов. Это используется для того, чтобы ИИ избегал повторов таких сочетаний слов. Например, в предложении «Я не уверен, смогу ли я когда-нибудь снова ходить с ним. Я не уверен, что…» есть повторяющееся сочетание из 3 слов «Я не уверен». Параметр является целым числом
* **Штраф за повторы** (repetition\_penalty). Это число в диапазоне от 0 до 1, где 1 означает отсутствие штрафа.
* **Top\_K.** Значение, которое отвечает за количество наиболее вероятных следующих слов. То есть, при включённом параметре, количество следующих слов уменьшается таким образом, что всем словам, кроме k самых вероятных ставится вероятность 0. Целое неотрицательное число.
* **Top\_P.** Вместо того чтобы делать выборку только из наиболее вероятных K слов, при выборке Top-p выбирается наименьший возможный набор слов, суммарная вероятность которых превышает вероятность p. Затем вероятностная масса перераспределяется между этим набором слов. Таким образом, размер набора слов (он же количество слов в наборе) может динамически увеличиваться и уменьшаться в зависимости от распределения вероятности следующего слова. Top-p может использоваться в комбинации с Top-K, что позволяет избежать слов с очень низкой вероятностью, обеспечивая при этом некоторый динамический отбор. Число в диапазоне от 0 до 1.
* **Температура** (temperature) – это параметр ИИ моделей, который регулирует случайность и, соответственно, креативность ответов. Это число в диапазоне от 0 до 1.

Температура 0 означает, что модель всегда будет выбирать слово с наибольшей вероятностью.

Более высокая температура означает, что модель может выбрать слово с несколько меньшей вероятностью, что приводит к большей вариативности, случайности и творчеству.

Очень высокая температура, таким образом, увеличивает риск "галлюцинаций", то есть ИИ начнет выбирать слова, которые не будут иметь смысла или будут не по теме.

Для задач преобразования (извлечение, стандартизация, преобразование форматов, исправление грамматики) предпочтительнее температура 0 или до 0,3.

Для задач написания текстов температуру следует выбирать выше, ближе к 0,5. Если вы хотите, чтобы ИИ был очень креативным (например, для маркетинговых или рекламных копий), используйте значения от 0,7 до 1.

Если вы хотите экспериментировать и быстро создавать множество вариантов, лучше использовать высокую температуру.

* **Штраф за длину** (length\_penalty). Экспоненциальный штраф к длине, используемый при лучевой генерации (beamsearch). Он применяется как экспонента к длине последовательности, которая, в свою очередь, используется для деления оценки последовательности. Поскольку оценка представляет собой логарифм вероятности последовательности (т.е. отрицательна), то length\_penalty > 0 способствует увеличению длины последовательностей, а length\_penalty < 0 - уменьшению длины последовательностей.

# Как происходит выбор лучшего варианта?

Для суммаризатора лучший вариант выбирается, используя метрику BertScore. Для рерайтера можно посмотреть лучший вариант по двум метрикам – BertScore и Classifier. Classifier – обученный sber классификатор на основе Roberta-large.