

Адаптация формальных текстов



СТРАТЕГИЯ
РАЗВИТИЯ
НИЖЕГОРОДСКОЙ
ОБЛАСТИ

Задача кейса



Разработка инструмента, который
будет помогать адаптировать
формальный канцелярский кейс на
простой язык, понятный человеку

Наше решение

Разработка специального расширения для браузера, которое благодаря методам МО позволит адаптировать формальный канцелярский текст на простой язык



Изменить текст

Простой

Ясный

Изменить страницу

Введите текст

Выберите файл

Преимущества

Retention ↑ 40%

Работает прямо в браузере, без необходимости переключаться на отдельный сайт.



Датасет

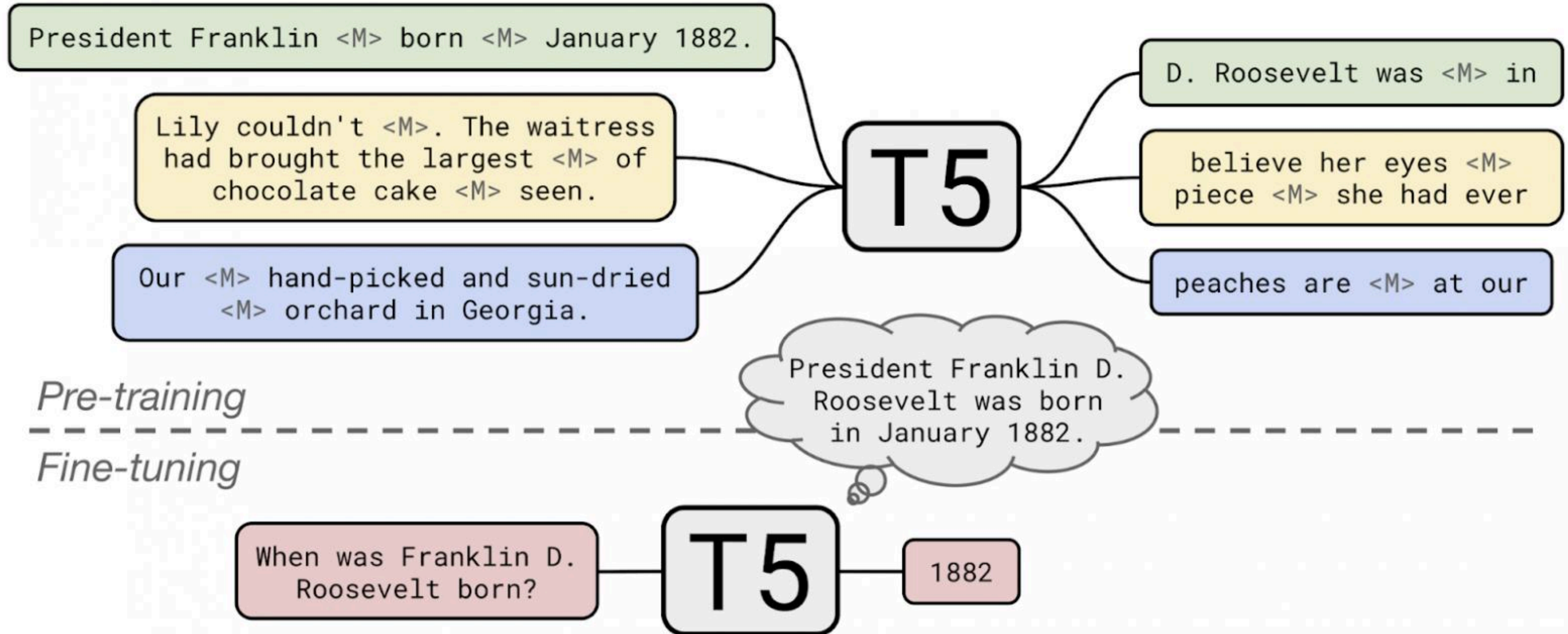
Официальные заявления Правительства РФ

2963 пары записей канцелярского и неформального текста

Документы с 31 декабря 2008 года по 28 ноября 2022 года



T5 model



Метрики

Cross-Entropy loss

$$\mathcal{L} = -\frac{1}{T} \sum_{t=1}^T \log p(y_t)$$

Чувствительна к производительности модели

Вычислительно затратно для больших словарей

Эпоха	Validation loss
1	2.41
2	2.17
3	1.82

Перспективы развития

Увеличение обучающего датасета

Дообучение модели mt5

Внедрение нового функционала

