

Rödl & Partner

ВМЕСТЕ К УСПЕХУ

GOOGLE TRANSLATE, DEEPL & CO.- ПРОБЛЕМЫ
МАШИННОГО ПЕРЕВОДА

ТОМАС ШТАЙНКЕ

Москва, 27 октября 2021 г.



СОДЕРЖАНИЕ

- Что такое машинный перевод
- Риски
- Примеры

ИСТОРИЯ

- с 1970-х гг. создание первых программ (на вьетнамский, с русского) для военных целей
- Первая коммерчески успешная система машинного перевода: SYSTRAN (использовалась Министерством обороны США), затем начала использоваться Европейской Комиссией
- 1996 г.: SYSTRAN - первый сервис машинного перевода в сети Интернет
- 2006 г.: появление онлайн-сервиса Google Translate



ПРИНЦИП ФУНКЦИОНИРОВАНИЯ



Изначально – чисто статистический принцип (дилемма: слово/словосочетание/предложение)



Анализ текстов в свободном доступе в сети Интернет



Сегодня – статистический принцип + нейронная сеть („Deep Learning“)





ЮРИДИЧЕСКИЕ РИСКИ

Microsoft Translator (Договор сервисного обслуживания): «Если Вы предоставляете Ваш контент другим лицам, то Вы в явной форме соглашаетесь с тем, что такие лица вправе использовать Ваш контент на безвозмездной основе во всех странах мира, хранить, записывать, копировать, отправлять, передавать, показывать Ваш контент и делиться им. [...]»



Положение о защите персональных данных
DeerL: «Обратите внимание, что использование нашего сервиса для перевода текстов, содержащих персональные данные любого рода, не допускается.»

Google: «Наши сервисы позволяют Вам загружать, отправлять, хранить, передавать и получать свой контент или обмениваться им. В этом случае Вы предоставляете компании Google право использовать, размещать, хранить, воспроизводить, изменять Ваш контент, создавать на его основе производные произведения (в т.ч. на основе перевода, корректировок или других изменений, осуществляемых нами для того, чтобы Ваш контент лучше функционировал в наших сервисах), передавать, публиковать, открыто воспроизводить, показывать и распространять Ваш контент.»

ЮРИДИЧЕСКИЕ РИСКИ

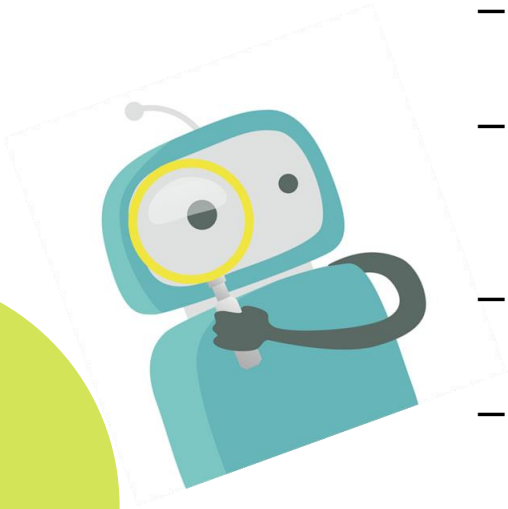
Сложно соблюдать
конфиденциальность

ПОСТОЯННАЯ
НЕОБХОДИМОСТЬ
АНОНИМИЗАЦИИ ДАННЫХ !!!!

Исходный текст
используется/сохраняется
(наиболее высокий риск при
использовании бесплатных
онлайн-версий)



ЛИНГВИСТИЧЕСКИЕ РИСКИ И РИСКИ, СВЯЗАННЫЕ С ПЕРЕДАЧЕЙ СМЫСЛА



- Пропуск слов, словосочетаний или целых предложений
- Использование разных терминов в одном и том же пункте (слово «Устав» может переводиться как charter, articles of association, bylaws и т.п.)
- Использование некорректных терминов в определенном контексте
- Дословный перевод, частично в сочетании с неправильным распознаванием смысловых оттенков (например, немецкое слово „Anhaltspunkt“ («подсказка», «намек на», дословно «точка остановки»), в DeepL как «brake point»)
- Система машинного перевода не распознает ошибки в исходном тексте



ПРИМЕРЫ



Israeli police arrested a Palestinian man last week after a Facebook post he made saying "good morning" in Arabic was mistranslated to read "attack them" in Hebrew

<https://www.bbc.com/news/world-middle-east-41714152>



GOOGLE TRANSLATOR

ВОЙТИ

Переводчик

С языка: английский

↔

На: русский

Перевести

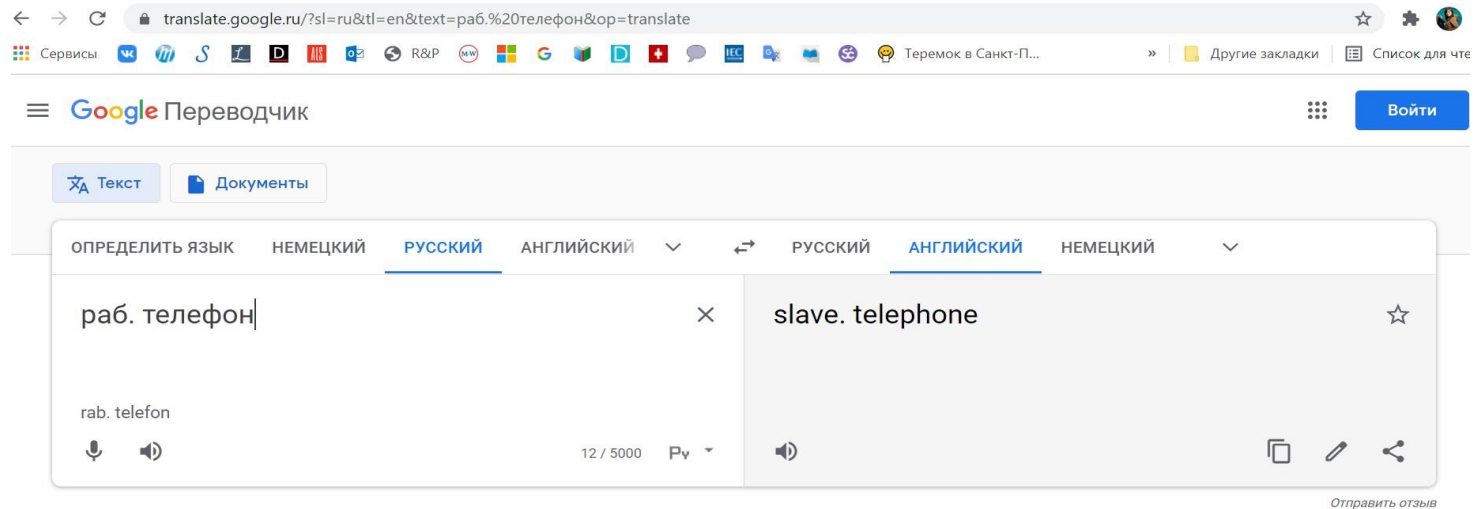
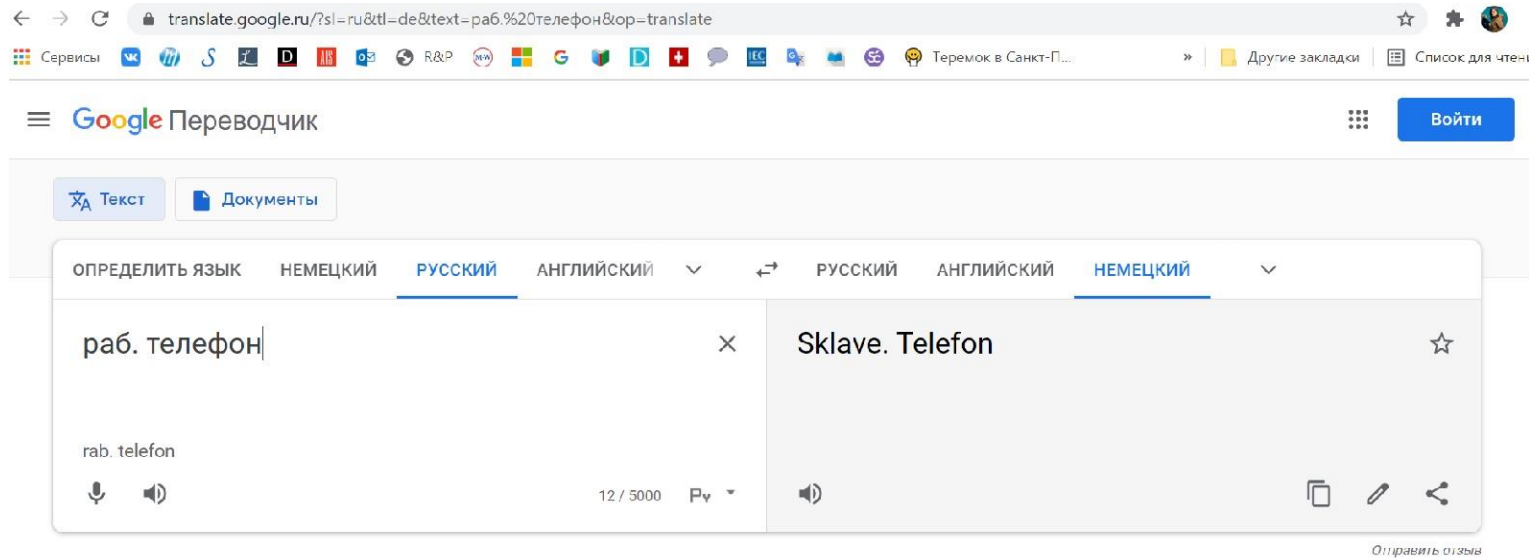
английскийрусскийнемецкийОпределить язык

Dear Customer:
Thank you for purchasing my products and the fastest speed with your payment.
Your goods has been shipped. It normally takes 20-30 days transit time. It may take even longer if there is abnormal weather condition or customs control.
Please don't worry about it.
I hope that this product is able to meet your needs and provide you with convenient. If you have any problems please do not hesitate to inform me about it. I am a reliable seller and would like to deal any problems with you. Wish you have a good mood.

русскийанглийскийукраинский

Уважаемые клиенты!
Благодарим Вас за покупку моей продукции и максимальной скорости с вашей компенсацией. Ваш товар будет отправлен. Это обычно занимает 20-30 дней, транзитное время. Это может занять еще больше, если есть ненормальное состояние погоды или таможенный контроль. Пожалуйста, не волнуйтесь об этом.
Надеюсь, что этот продукт способен удовлетворить ваши потребности и предоставить Вам удобно. Если у вас возникли проблемы, пожалуйста, не стесняйтесь сообщить мне об этом. Я надежного продавца и хотел бы иметь дело все проблемы с вами. Жаль, что у вас хорошее настроение.

GOOGLE TRANSLATOR



КАРТИНА НЕФТЬЮ, НЕ ПЕРЕЖИВАЙТЕ!

The Brook

Paul Cezanne

Date: c.1898 - 1900

Style: Post-Impressionism

Genre: landscape

Media: oil

Location: Cleveland Museum of Art
(CMA), Cleveland, OH, US

Брук

Пол Сезанн, Пол

Дата: с. 1898-1900

Стиль: пост-импрессионизм

Жанр: пейзаж

СМИ: нефть

Местонахождение: Музей
искусства Кливленда (CMA),
Кливленд, OH, US



МАРК ШАГАЛ И РАШН ОРИДЖИНАЛ





+7 (495) 154-78-82
ОСТАВИТЬ ЗАЯВКУ

SHAGAL В ГОРОДЕ

Новый проект расположен
в Даниловском районе
прямо на набережной Марка Шагала.

|

Everything is near!



ru.m.wikipedia.org

Шагал, Марк Захарович

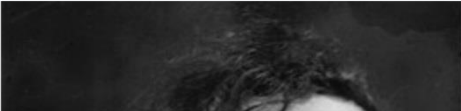
🔍 ☆ ✎

Запрос «Шагал» перенаправляется сюда; см. также [другие значения](#).

Марк Заха́рович (Моисе́й Ха́цкелевич) Шага́л (фр. *Marc Chagall*, идиш מאַרק שאַגאַל; 6 июля^[5] или 7 июля^[6] 1887, Лиозно, Витебская губерния, Российская империя — 28 марта 1985, Сен-Поль-де-Ванс, Прованс, Франция) — русский и французский художник еврейского происхождения. Помимо [графики](#) и [живописи](#) занимался также [сценографией](#), писал стихи на [идише](#). Один из самых известных представителей художественного [авангарда](#) XX века^{[1][7]}.

Марк Захарович Шагал

Моисей Хацкелевич Шагал^[1]







КЛАССИКА ЖАНРА



КОНТАКТНОЕ ЛИЦО



ТОМАС ШТАЙНКЕ

Ассоциированный партнер
Дипломированный устный и письменный переводчик
Руководитель отдела переводов СНГ

Тел. +7 495 933 5120
thomas.steinke@roedl.com

Rödl & Partner
Бизнес-центр «ЛеФорт»
Электрозаводская ул., 27, стр. 2
107023 Москва
Россия