



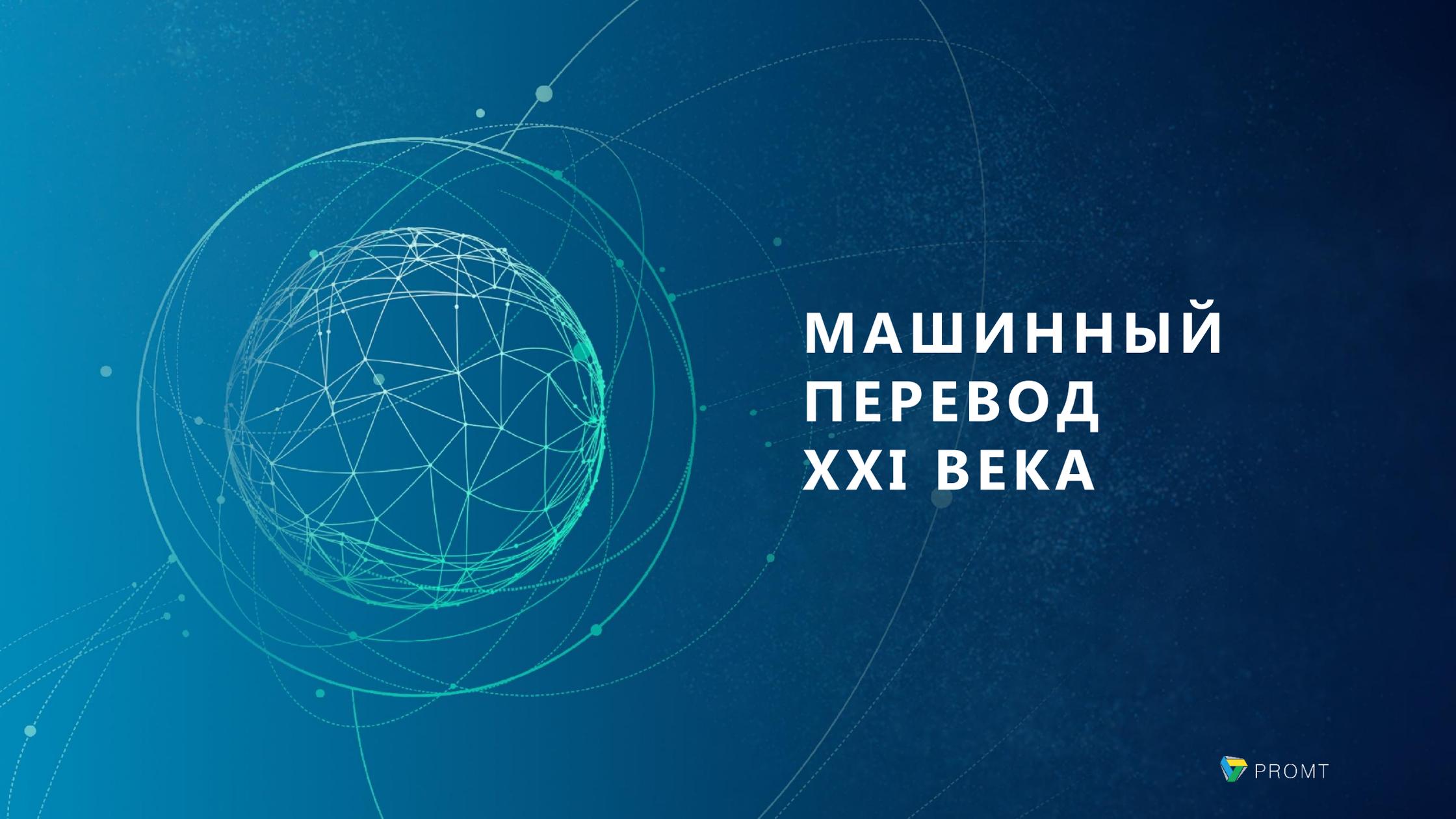
Машинный перевод: РЕФОРМАТОР МИРОВОЙ ЭКОНОМИКИ

Юлия Епифанцева
PROMT



Перевод как стратегическая задача

- Евросоюз. 28 языков. Информация должна быть представлена на всех языках ЕС
- Программы ЕС по развитию науки и технологий Horizon 2020 (бюджет на 7 лет 80 миллиардов евро)
- Выручка каждой из ста крупнейших переводческих компаний в мире составляет от 10 до 700 миллионов долларов
- По данным Global Market Insights, Inc. к 2021 году объём мирового рынка лингвистических услуг составит 56 миллиардов долларов



МАШИННЫЙ ПЕРЕВОД XXI ВЕКА

ИИ: задачи, решаемые человеком – может решать машина

Машинный перевод – одна из самых успешных и широко распространенных реализаций Искусственного Интеллекта (ИИ)



Распознавание речи



Классификация текстовых сообщений в потоке



Распознавание образов (предметов, явлений, лиц)



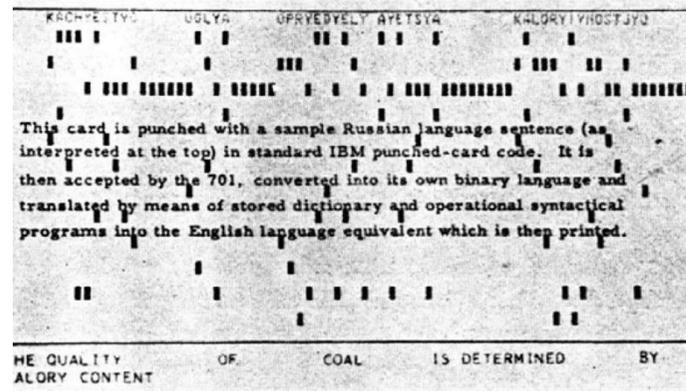
Идентификация людей по голосу



Перевод текста

7 января 1954 года
Georgetown University & IBM

Русско-английский переводчик:
6 грамматических правил, 250 лексических единиц; было переведено 200 фраз



Почему машинная обработка естественного языка сложная задача



МНОГООБРАЗИЕ КОНТЕНТА

Заголовки и большие тексты, распознанная устная речь, терминология из разных отраслей науки и бизнеса, деловой, научный стиль и сленг представляют собой «мини-языки» внутри большого языка

FOLHFY
BVDETFJ
ESZXVD

СРЕДНЯЯ ДЛИНА СЛОВ В РАЗНЫХ ЯЗЫКАХ

Увеличение или уменьшение количества символов в тексте перевода по сравнению с оригиналом влияет на сохранение форматирования документа при переводе.



АСИММЕТРИЯ ЯЗЫКОВ

Разные языки по-разному описывают мир и используют разные подходы в морфологии, грамматике, синтаксисе, семантике.



ДАННЫЕ ДЛЯ ОБУЧЕНИЯ

Сложности в поиске и подборе данных для обучения и настройки систем на перевод под конкретные бизнес-задачи

Технологии машинного перевода в мире

КРАТКАЯ ИСТОРИЯ МАШИННОГО ПЕРЕВОДА:

Rule-Based MT с 1950 гг.

Машинный акцент
в переводе,
предсказуемый
результат

ОБУЧЕНИЕ ЧЕРЕЗ
СЛОВАРИ

Statistical MT с 2000 гг.

Более гладкий
перевод чем в RBMT,
но непредсказуемый
результат

ОБУЧЕНИЕ ЧЕРЕЗ
ПАРАЛЛЕЛЬНЫЕ ДАННЫЕ

Neural MT с 2015 гг.

Точный и гладкий перевод,
результат на уровне
профессионального
переводчика

ОБУЧЕНИЕ ЧЕРЕЗ
ПАРАЛЛЕЛЬНЫЕ ДАННЫЕ

Данные в цифровой экономике – это новая нефть.

Драйверы перевода на основе нейронных сетей



Рост
вычислительных
мощностей



Рост
цифрового
контента



Развитие
нейросетевых
технологий

WMT Conference

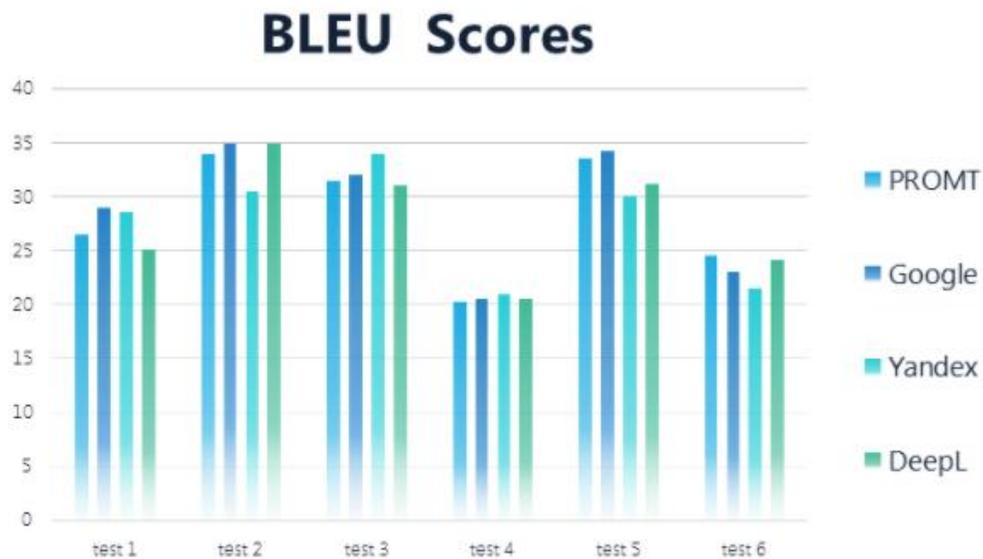
- Главное мероприятие по МП в мире
<https://www.aclweb.org/anthology/venues/wmt/>
- WMT 2020 состоится 18 ноября в онлайн-формате

English–Russian

#	score	range	system
1	1.015	1	PROMT-RULE
2	0.521	2	ONLINE-G
3	0.217	3	ONLINE-B
4	0.122	4-5	LIMSI-NCODE
	0.075	4-5	ONLINE-A
5	0.014	6	UEDIN-JHU
6	-0.138	7	UEDIN-SYNTAX
7	-0.276	8	USAAR-GACHA
8	-0.333	9	USAAR-GACHA
9	-1.218	10	ONLINE-F

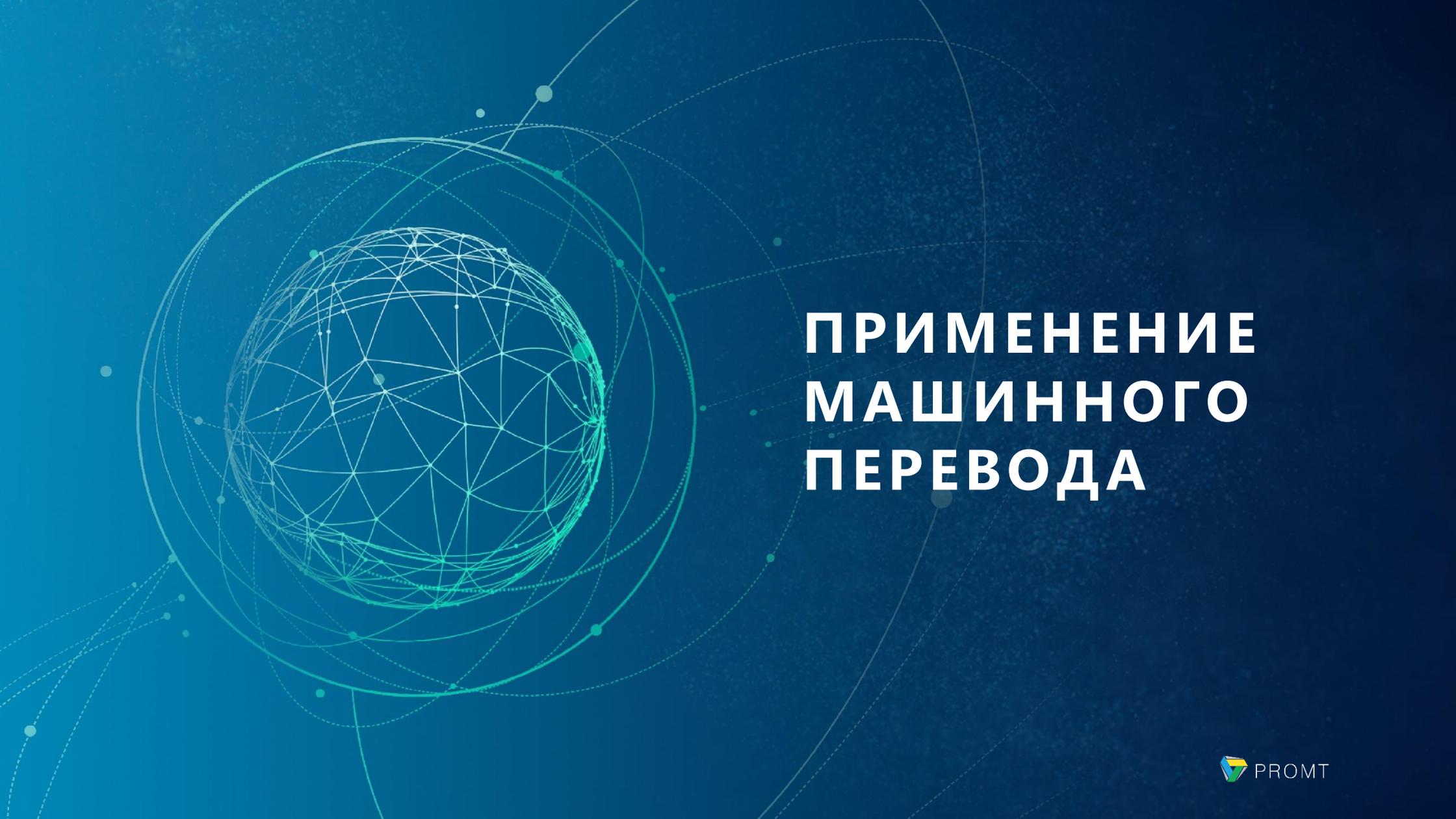
PROMT Neural, Google, Yandex, DeepL

Тестирование на текстах общей тематики, ER



Экспертная оценка

- перевод легко читается
- корректная структура предложения
- нет ошибок согласования
- перевод требует незначительного или вообще не требует редактирования



ПРИМЕНЕНИЕ МАШИННОГО ПЕРЕВОДА

Сценарии применения МП

СЛУЖБЫ ПОДДЕРЖКИ

для оперативного ответа на запросы

E-COMMERCE

локализация контента для покупателей

КОРПОРАТИВНАЯ ПЕРЕПИСКА

общение внутри глобальных компаний

ДОКУМЕНТАЦИЯ К ПРОДУКЦИИ

документация к ПО,
техническая документация

ЛОКАЛИЗАЦИЯ САЙТОВ

перевод контента сайтов на лету

ПОИСК ИНФОРМАЦИИ

поиск информации в мультиязычных БД,
на сервисах

ИНТЕГРАЦИЯ В СЕРВИСЫ И ПРИЛОЖЕНИЯ

как дополнение к основным функциям

ОБРАЗОВАТЕЛЬНЫЕ СЕРВИСЫ

расширение возможностей для сервисов
по изучению иностранных языков

Взаимодействие пользователей в онлайн-пространстве

Крупнейший в мире туристический сайт (переведен на 21 язык)

Основной контент – миллионы отзывов пользователей со всего мира (более 385 миллионов отзывов)

Машинный перевод с английского на русский

60 000+ запросов и 450 000+ слов в день



Обеспечение работы служб поддержки и общения с клиентами



amadeus

Одна из крупнейших в мире
глобальных систем онлайн-
дистрибуции билетов

Условия бронирования формируются
в режиме реального времени
на английском языке

Более **5 000** пользователей
(туристических агентств в России)

Более **6 000** запросов на перевод или
более **4 млн** слов в неделю

Международное сотрудничество



Перевод внутренней документации с английского, немецкого, испанского и др.

Пользователи в филиалах по всему миру (Австралия, Северная и Южная Америка, ЮАР, Финляндия, Болгария и др.)

Быстрый перевод для понимания деловой переписки, юридических и бухгалтерских документов

Экономия времени и ресурсов - перевод у бизнес-пользователей в режиме реального времени

Оптимизация затрат на профессиональный перевод



Обновление документации на сайтах
и в новых версиях продуктов

Перевод в короткие сроки
с английского на 30 языков

С машинным переводом рост
производительности
переводчиков на 15+%

Изучение иностранных языков

Опрос пользователей сервиса PROMT.One (www.translate.ru) и в мобильных приложениях PROMT

Ноябрь-декабрь 2016 г.
Участники: 8598 человек

Для чего вы используете переводчики PROMT?

Изучаю иностранный язык самостоятельно или на курсах

39,9%

Готовлю домашние задания в школе, институте, аспирантуре

25,3%

Помогаю детям с домашними заданиями

16,2%

Онлайн-образование (выбор курсов, знакомство с материалами)

12,3%



Сколько книг переводится на PROMT.One в день?

Всего переведено за
1 год = 69 млрд слов
В среднем за
1 день = 190 млн слов

Для перевода **190 млн слов**
нужно:
31 717 дней = 87 лет
работы без перерыва 1 переводчика;
76 120 переводчиков,
чтобы перевести все за 1 день



COVID. Сервисы для врачей

Российский сервис для врачей, где интегрированы технологии перевода PROMT для перевода статей и другой медицинской и фармакологической информации из зарубежных источников.

В марте и апреле 2020 общее количество запросов на перевод увеличилось в 160 раз, а объемы перевода – в 55 раз!

5 причин задуматься о МП

- Необходимость перевода с и на иностранные языки никогда не исчезнет
- Технологии будут все больше и больше использоваться для дополнения и расширения работы профессиональных переводчиков
- Нейронный машинный перевод — это инновация и революция в области ИИ
- Актуальный вызов — это не только перевод текста, но и перевод документа
- Вопросы безопасности в ИТ не утратят своей значимости

Спасибо за внимание!

julia.epiphantseva@promt.ru

+7 (495) 580 48 48

promt.ru