



НОВЫЕ ЛИНГВИСТИЧЕСКИЕ
ТЕХНОЛОГИИ

Локализация новостного сайта с помощью машинного перевода

Локализация новостного портала TAdviser

Юлия Епифанцева

Директор по развитию PROMT

О компании

PROMT – ведущий российский разработчик лингвистических IT-решений для корпораций и частных пользователей. Компания предоставляет весь спектр лингвистических IT-решений: универсальные и специализированные системы машинного перевода, словари, API переводчика, системы анализа текстов, решения для работы с Big Data, мобильные приложения для всех платформ, CAT-системы.

- Более 30 лет успешной работы в РФ и в мире
- Лучшие эксперты в области МП для корпоративного сегмента в России с 1991г.
- Владелец первого в Рунете бесплатного сервиса [онлайн-перевода PROMT.One](#)



Направления разработок компании

ПО для машинного перевода для компаний

Особенности ПО

- Технологии на основе нейронных сетей для перевод текстов и документов
- Поддержка 50+ языков
- Возможность повышать стартовое качество перевода за счет обучения нейронных моделей и создания словарей на основе глоссариев
- On-premise решения, т.е. установка на серверах организации и гарантия безопасности и конфиденциальности переводимых данных
- 100% российская разработка, подтвержденная сертификатами совместимость с ОС на базе Linux, продукты в реестре отечественного ПО.

Успешные проекты

СИБУР

Внедрение корпоративной системы перевода. Система доступна всем пользователям Холдинга – более 20 тыс. сотрудников.



Банк России

Внедрение серверного решения PROMT и его кастомизация на основе корпоративных глоссариев и ранее выполненных переводов. Улучшение качества перевода на 25%. 1500 уникальных пользователей.

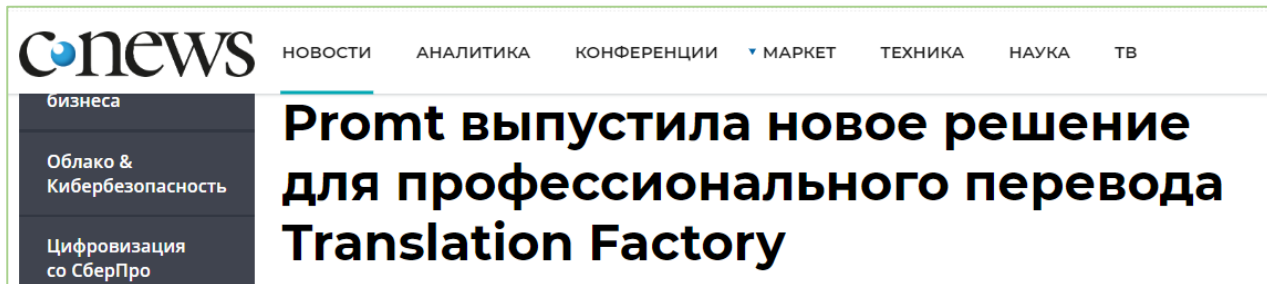
Сергей Кожевников, Консультант Управления информационно-библиотечного обеспечения Центрального Банка России

«Сразу после внедрения мы видим рост заявок на перевод на английский беспрецедентно большого объема. Совершенно точно можно сказать, что без машинного перевода выполнить такой объем в поставленный срок было бы просто невозможно».

Направления разработок компании

ПО для профессионального перевода

Первая российская многопользовательская CAT-система (computer added translation) для конфиденциального перевода в локальной сети компаний и организаций – PROMT Translation Factory (*релиз осень 2023*)



Платформа позволяет поддерживать основные бизнес-процессы в области профессионального перевода: от управления переводческими проектами, внедрением справочных ресурсов и баз данных до интеграции технологий искусственного интеллекта для машинного перевода и инструментов кастомизации решения на своих данных.

Может заменить

- ✓ Trados
- ✓ Phrase
- ✓ MemoQ
- ✓ Smartcat

Огромный интерес от компаний банковского сектора, атомной энергетики и нефтегазовой отрасли

Направления разработок компании

Сервис бесплатного перевода, мобильные приложения

- Первый бесплатный **сервис онлайн-перевода Рунета** (1998) – PROMT.One (Translate.RU)
- Нейронные технологии перевода, 20+ языков
- **Контекстный словарь** , чтобы просматривать перевод миллионов слов и выражений и изучать их употребление на реальных примерах.
- **Сервис Склонение и спряжение** для 7 языков
- **Мобильные приложения** для iOS и Android и бот-переводчик для Telegram
- **API переводчика** (Почта MailRu, SpanishDict)



- ✓ Интеграция PROMT.One API в почтовый сервис (с 2008 года)
- ✓ Автоматический перевод входящих и исходящих писем в один клик.
- ✓ Постоянные обновления языковых моделей PROMT.One API

Машинный перевод в локализации

Опыт PROMT в локализационных проектах

amadeus

Автоматический перевод условий бронирования в системе Amadeus

KASPERSKY

Перевод документации к продуктам

 **tripadvisor**

Машинный перевод пользовательских отзывов

- ✓ Машинный перевод с постредактированием
- ✓ Машинный перевод без постредактирования
- ✓ Кастомизация машинного перевода
- ✓ Метрики для автоматической оценки качества перевода

АНГЛОЯЗЫЧНЫЙ ПОРТАЛ НА ОСНОВЕ НЕЙРОННОГО МАШИННОГО ПЕРЕВОДА

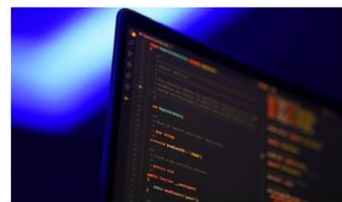
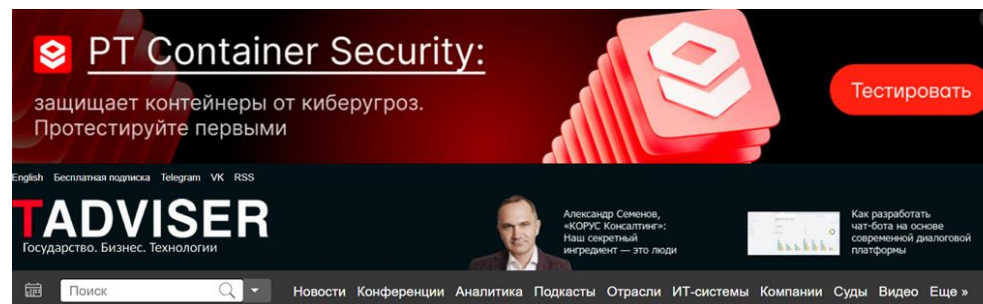
Информационный портал Tadviser.ru




TAdviser (от англ. *Technologies* – технологии и *Adviser* – советник, консультант, наставник) - российский интернет-портал и аналитическое агентство.

Крупнейший в России портал по теме корпоративной информатизации

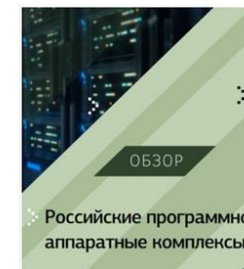
Ведущий организатор мероприятий по темам корпоративной информатизации



Почему ИТ-рынок России вырос, несмотря на санкции США. Мнение ВШЭ  436,5



Максут Шадаев на TAdviser Summit — о вызовах в цифровизации госсектора, давлении китайских



Задачи и требования со стороны заказчика

В России сильная индустрия информационных технологий, но если информация доступна только на русском языке, то это ограничивает доступ зарубежной аудитории.

По данным Международного союза электросвязи (ITU), в интернет выходят 3,9 млрд человек или 51,2% населения планеты, из которых как минимум половина делают поисковые запросы на латинице, поэтому важно представить информацию прежде всего на английском языке, а в перспективе и на других языках.

Основные требования со стороны заказчика:

- качество перевода с русского на английский без постредактирования, достаточное для понимания смысла, автоматический перевод всего контента сайта,
- повторный перевод статей и новостей, если в них вносились правки на русском языке,
- возможность оперативной точечной коррекции перевода.

Первый этап

Анализ данных и планирование разработки

Первый этап 2018 - осень 2020 гг.

- Анализ русскоязычного контента TAdviser (4 типа текста - новость, интервью, биографическая карточка, аналитика),
- Настройка для разных типов текста (новость, интервью, биографическая карточка) для повышения качества машинного перевода,
- Особое внимание к переводу имен собственных, для настройки ТА предоставил списки компания и персоналий,
- Тестирование качества перевода

*На первом этапе использовалась технология Rule-based Machine Translation
Лингвистические базы данных (словари, списки имен) + Специальные алгоритмы для перевода
любых имен собственных, в том числе отсутствующих в лингвистических базах данных.*

Первый этап

Перевод имен собственных

Первый этап 2018 - осень 2020 гг.

Создание словарь с именами собственными:

- Словарь компаний (2620 названий)
- Словарь персон (2714 имен)
- Словарь названий ПО (373 наименований)
- Далее были дополнения
 - должности,
 - министерства,
 - законы и подзаконные акты
 - названия ИТ продуктов компаний

Первый этап

Разработка и интеграция

Первый этап 2018 - осень 2020гг.

- ✓ Разработка TAdviser
 - HTML страница на русском
 - Веб-сайт на русском (PHP-модули)
 - База данных с текстами в формате Wikitext на русском
 - **PHP-компонент для запросов на перевод**
 - HTML-страница на английском языке
 - Веб-сайт на английском (PHP-модули)
 - База данных с текстами в формате Wikitext на английском

Wikitext – это документ в wiki-разметке (контент, разметка, метатэги)

- ✓ Разработка PROMT
 - Веб-сервис для перевода wiki-строк
 - Сервис машинного перевода, модуль МП для русско-английского
 - Настроенный профиль перевода

Первый этап Wikitext

Первый этап 2018 - осень 2020гг.

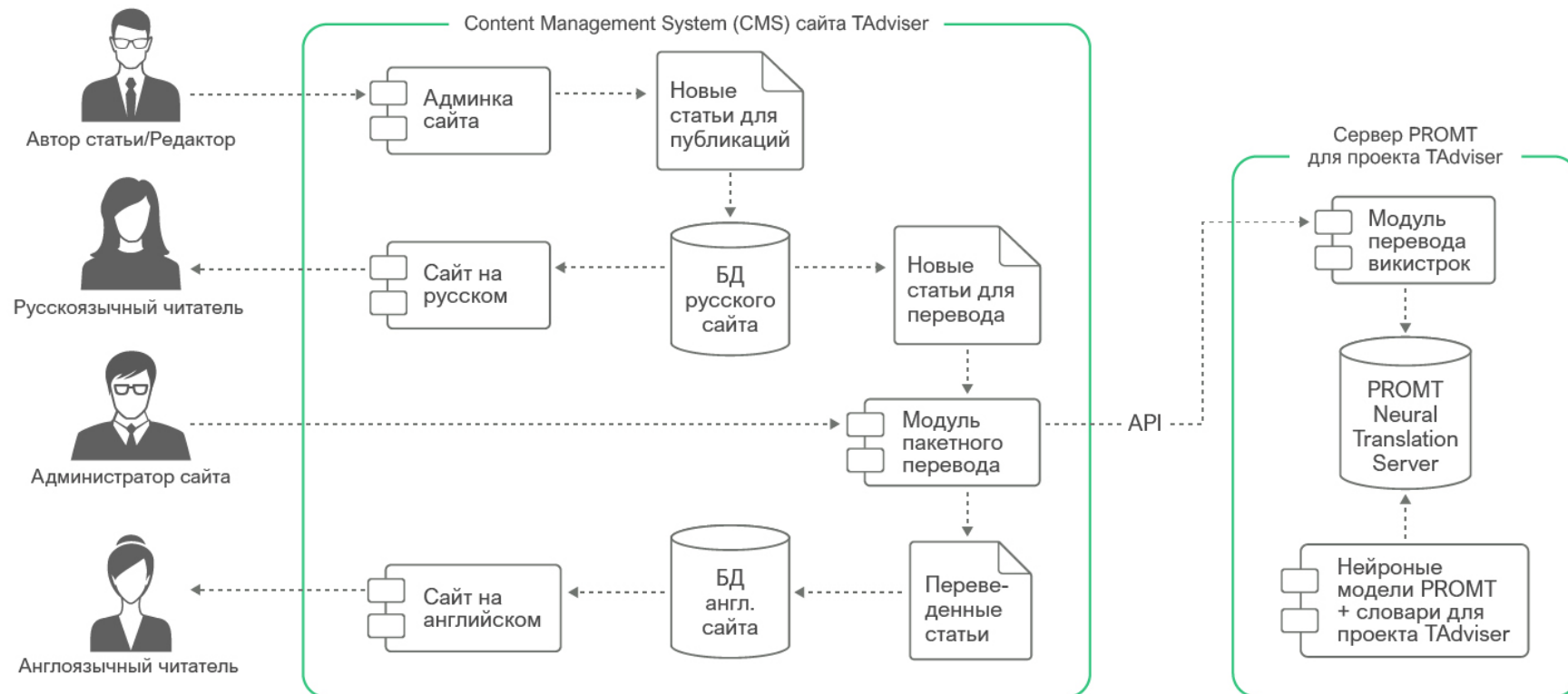
Wikitext – это документ в wiki-разметке (контент, разметка, метатэги)

По мнению экспертов, публичные облачные сервисы становятся не просто альтернативой собственной инфраструктуре предприятия или ИТ-сервисам предприятия, организации все чаще используют публичные облака в качестве основы для доступа к новым технологиям, включая контейнерную виртуализацию, распределенные базы данных, аналитику больших данных, искусственный интеллект, интернет вещей. Это

По мнению экспертов, публичные облачные сервисы становятся не просто альтернативой собственной инфраструктуре предприятия или ИТ-сервисам предприятия, организации все чаще используют публичные облака в качестве основы для доступа к новым технологиям, включая [[ИТ-контейнеры|контейнерную]] [[виртуализацию]], распределенные [[БД|базы данных]], [[ВІ|аналитику]] [[больших данных]], [[искусственный интеллект]], [[интернет вещей]].

Первый этап

Схема интеграции



Второй этап

Переход на нейронный машинный перевод

Второй этап октябрь 2020-2021

Переход на технологию перевода на основе нейронных сетей
PROMT Neural Machine Translation:

- тестирование качества перевода по новой технологии,
- использование новых инструментов для настройки,
- выбор новой инфраструктуры для установки программного обеспечения и обеспечения требуемой скорости перевода контента

Сервер-перевода
CPU Intel Core i7 3770 3.4 GHz, 16GB RAM, Windows Server
2016

Второй этап

Кастомизация через нейронный словарь

Нейронный словарь для переводчиков PROMT

PROMT Neural Dictionary – специальная технология нейронного словаря, реализованная в нейронных продуктах системы перевода PROMT и доступная для Windows- и Linux-систем. Настройка через пользовательский интерфейс в ручном или в автоматическом режиме.



SMART Neural Dictionary

- учитывает контекст
- термины в переводе в нужном роде, числе и падеже
- Англо-русско-английский, немецко-англо-русский и др.

SIMPLE Neural Dictionary

- без учета контекста
- термины остаются в той же форме, что и в словаре
- 10+ языков, 45+ языковых пар

Второй этап

Smart&Simple Neural Dictionary

Исходный текст

Smart Neural Dictionary

The **cable ladder** is preferred to ensure cooling cables.

Перевод без глоссария

Кабельная лестница предпочтительна для обеспечения охлаждающих кабелей.

Перевод с глоссарием

Предпочтение отдается **лестничному лотку** для обеспечения охлаждающих кабелей.

Simple Neural Dictionary

Tenir compte des **consignes** techniques.

Beachten Sie die technischen **Anweisungen**.

Beachten Sie den technische **Vorgaben**.
Правильно: ... die technischen Vorgaben

Третий этап

Третий этап 2022- настоящее время

- Обновления нейронной модели (1 раз в год)
- Пополнение нейронных словарей по технологии PROMT Smart Neural Dictionary (по запросу)



Выпущена первая в России открытая библиотека для создания графовых нейросетей

Как пояснили МТС, графовые нейросети отличаются от обычных моделей машинного обучения тем, что учитывают любые взаимосвязи в данных и обобщают все предыдущие подходы. Поэтому нейросеть получается «более гибкой и вариативной». Например, чат-бот сможет анализировать каждое слово в контексте предыдущих, выдавая



The first open library in Russia for creating graph neural networks has been released

As MTS explained, graph neural networks differ from conventional machine learning models in that they take into account any relationships in the data and summarize all previous approaches. Therefore, the neural network is "more flexible and variable." For example, a chatbot will be able to analyze each word in the context of the previous ones, giving

Объемы переводимого контента

К моменту запуска система перевела с русского на английский более 206 тысяч статей из базы знаний TAdviser, в том числе более 40 000 досье компаний, 17 000 описаний ИТ-решений и 47 000 описаний ИТ-проектов, а также большое число статей о технологических рынках России и других стран, сотни интервью с экспертами из отрасли информационных технологий.

В настоящий момент ежемесячно автоматически переводится 60 млн символов (2023-100 млн), это более 33 000 страниц текста формата А4. Если представить, что один переводчик может перевести в месяц около 200 страниц текста, то для перевода всего месячного потока понадобилось бы не менее 150 переводчиков.


Результаты & Сложности проекта

- Всего на проект было потрачено – около 2 700 ч/ч с 2018 г.
- Команда TAdviser - 2 человека (менеджер, веб-разработчик)
Команда PROMT - до 5 человек (менеджер проекта, разработчик модуля парсинга, тестер, лингвист-разработчик, администратор облачного сервиса) + готовая система МП и нейронная модель перевода.
- Перевод без постредактирования/модерации человеком.
- Необходимость специальной технической обработки контента (разработка алгоритма викитекста, включающего в себя пре- и пост-процессинг викиссылок внутри предложений).
- Необходимость оперативной точечной коррекции переводов имен собственных по запросу заказчика (без переобучения моделей).

Tadviser.com

- англоязычная версия сайта формируется автоматически
- контент переводится фактически в режиме реального времени
- кастомизированное решения (настройка перевода имен собственных - названия компаний, организаций) и возможность дополнительной точечной настройки



Yandex.Eda will pay compensation to users for leakage of their data  2300

По данным сервиса Similarweb.com на tadviser.com ежемесячно фиксируется от 90 000 до 130 000 визитов. География пользователей портала: США, Россия, Великобритания, Турция и другие.



Over 25 billion rubles will be allocated for the development of state cybersecurity systems until 2030

Англоязычный портал на основе нейронного автоматического перевода

ЗАКАЗЧИК

TAdviser

РУКОВОДИТЕЛЬ ПРОЕКТА
СО СТОРОНЫ ЗАКАЗЧИКА



Александр Левашов
Главный редактор

ИТ-ПОСТАВЩИК

ООО ПРОМТ

ГОД ЗАВЕРШЕНИЯ ПРОЕКТА

2021

СРОКИ ВЫПОЛНЕНИЯ
ПРОЕКТА

Май, 2018 - Октябрь, 2021

МАСШТАБ ПРОЕКТА

2300 человеко-часов

News > Компания PROMT и...

Компания PROMT и интернет-портал TAdviser принимают участие в конкурсе «Проект Года» от Global CIO



20 December 2021 Company's news PROMT

«Проект года» — это ежегодный конкурс проектов в сфере ИТ-технологий. Конкурс создан с целью объединения практических знаний в области информационных технологий.

Организатором конкурса является портал Global CIO | DigitalExperts — это профессиональное сообщество лидеров цифровой трансформации. В него входит свыше 10 000 участников. Цели сообщества: развитие цифровизации в нашей стране, сохранение практического опыта в единой базе знаний и поддержка общения внутри ИТ-сообщества.

2021/03/02 11:00:00

TAdviser.com - первое в России издание, работающее на основе искусственного интеллекта, переведено на нейросетевые технологии

TAdviser.com - англоязычная версия российского издания TAdviser.ru, формируется автоматически с использованием решений машинного перевода компании PROMT. Это первый проект подобного рода в России и, вероятно, в мире.



НОВЫЕ ЛИНГВИСТИЧЕСКИЕ
ТЕХНОЛОГИИ

Юлия Епифанцева

Директор по развитию бизнеса

Julia.Epiphantseva@prompt.ru

www.prompt.ru