






**Искусственный интеллект  
на основе нейросетей для  
работы с документацией  
на иностранных языках  
и интеллектуальный анализ  
неструктурированных  
данных**



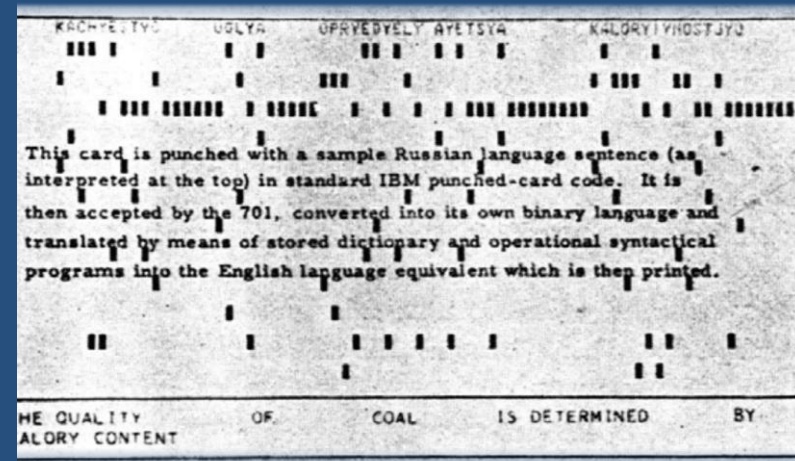
# Определение ИИ: задачи, решаемые человеком – может решать машина

Автоматический 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



# Развитие машинного перевода

## Аналитический МП (АМТ)

- + Надежность
- + Обучение
- + Скорость
- + Минимальные требования
- Машинный акцент
- Стоимость разработки

## Статистический МП (SMT)

- + Гладкий
- + Стоимость
- + Обучение (параллельные данные)
- +/- Надежность
- Скорость
- Требования к аппаратным ресурсам
- Зависимость результата от объема и качества обучающих данных

# Нейронный МП (NMT)

- + Качество, близкое к переводу, выполненному человеком
- + Возможность перевода контента, который отсутствует в обучающих данных
- + Стоимость разработки
- + Обучение специальной терминологии

- Надежность
- Отсутствие возможности настройки без параллельных данных
- Особые требования к аппаратным ресурсам
- Зависимость результата от объема и качества обучающих данных

# Что делать? Синергия всех технологий

ОБЪЕДИНЕНИЕ РАЗНЫХ ТЕХНОЛОГИЙ В РАМКАХ ОДНОГО РЕШЕНИЯ

- **Выбор от конкретной задачи**

Новые языковые пары и есть данные? → статистический или нейронный подход

Ограничены аппаратные ресурсы? → аналитический подход

- **Суперпозиция технологий или гибридный подход**

Статистическое или нейронное постредактирование

- **Классификация контента по тематике, структуре и особенностям**



# Современное состояние технологий машинного перевода PROMT на основе NMT

## «Чистый» NMT

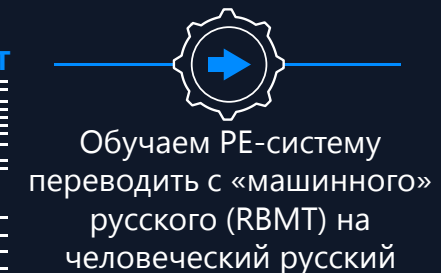
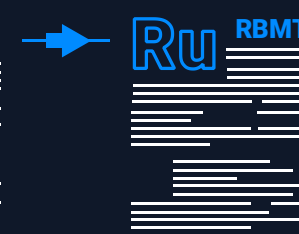
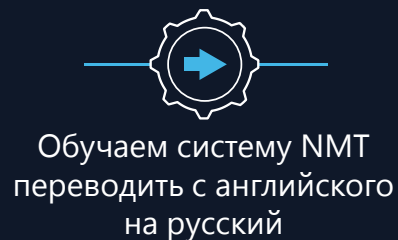


## Гибридный МП (RBMT+нейросетевое постредактирование)



### СИСТЕМЫ

### ОБУЧЕНИЕ



# 1 клик мыши отделяет от утечки информации

## УГРОЗА!

Утечка информации, потеря конфиденциальности которой – критична

Перевод через онлайн-сервис создает только иллюзию безопасности



# Онлайн-сервис как канал утечки данных

2017 год. Скандал вокруг известной компании Statoil. Тысячи конфиденциальных документов стали общедоступными.

Они были переведены американским сервисом онлайн-перевода **translate.com**, и проиндексированы поисковыми системами

**Любой загруженный контент будет проанализирован с неочевидными последствиями** как для сотрудника, так и **для организации**, в которой он работает

Бесплатная услуга = Предоставление своих данных в свободное пользование сервиса





# Условия использования сервисов онлайн-перевода



...не может и не гарантирует, что любая информация, предоставленная нам вами, не станет общедоступной ни при каких обстоятельствах.

Вы должны понимать, что вся информация, представленная на веб-сайте, потенциально может быть общедоступной.



- When you upload, submit, store, send or receive content to or through our Services, you give Google (and those we work with) a worldwide license to use, host, store, reproduce, modify, create derivative works (such as those resulting from translations, adaptations or other changes we make so that your content works better with our Services),
- communicate, publish, publicly perform, publicly display and distribute
- such content.

# Какую информацию можно получить из отрывков текста?

## ЧАСТОТНЫЙ И ЛИНГВИСТИЧЕСКИЙ АНАЛИЗ

свободно покажет частоту употребления тех или иных слов в связке с:

- событиями
- именами
- датами
- и любыми другими данными

## КОНТЕНТ ФИЛЬТРУЕТСЯ ПО IP-АДРЕСУ И ID УСТРОЙСТВА

## АНОНИМАЙЗЕР НЕ ПОВЛИЯЕТ НА РЕЗУЛЬТАТЫ

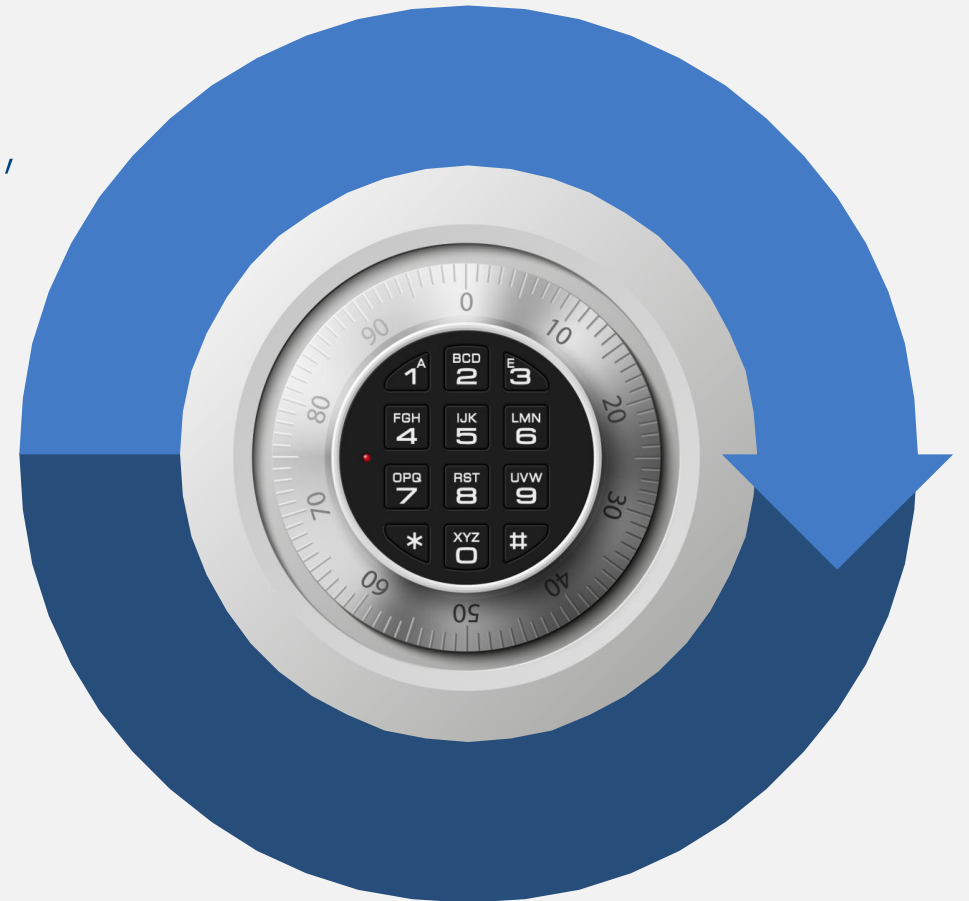


# Экспертное мнение

PROMT как компания-эксперт в области технологий перевода хорошо осознает риски, возникающие при использовании облачных решений при переводе конфиденциальной информации.

**100%** БЕЗОПАСНОСТЬ  
КОНТЕНТА

Обеспечивает только внутренний сервис автоматического перевода, не соприкасающийся с внешней средой!



# Альтернативное решение для безопасного перевода

## PROMT Translation Server:

- Устанавливается во внутреннюю сеть организации
- Переводит в офлайн-режиме
- Интегрируется в ИАС
- Содержит конфиденциальный режим перевода документов (ограничен доступ даже для администратора сервиса)



**Полностью исключает попадание информации к третьим лицам!**

# Передовые технологии PROMT: преимущества

- Аналитические, статистические, нейронные и гибридные технологии перевода
- Offline перевод в закрытом контуре
- Высокое качество перевода
- Перевод документов с сохранением форматирования
- Широкие возможности гибкой настройки
- API/SDK для интеграции в другие программные комплексы и приложения
- Использование технологий отечественного производителя
- Решения включены в Реестр российского ПО
- Поддержка и команда разработки в РФ





# Машинный перевод: задачи для ВПК

Использование  
конфиденциального  
перевода  
(без переводчика)  
на переговорах

Перевод инструкций  
к используемому  
зарубежному  
оборудованию, для  
усовершенствования  
разработок

Перевод  
переписки,  
исключая риск  
утечки данных

Мониторинг  
СМИ и  
соцсетей



# 30 лет успешной работы с конфиденциальной информацией

Компания PROMT с 1991 года разрабатывает решения\*, обеспечивающие полную конфиденциальность при работе с любой информацией



\* -Все решения PROMT внесены в Единый реестр российских программ для ЭВМ и БД

# От Big Data к Smart Data

Анализ неструктурированных данных на разных языках – еще одна задача для искусственного интеллекта.

Big Data – ключевой драйвер развития цифровой экономики.

Оперативная реакция на изменения возможна только с использованием **Smart Data**.



# Что используется для аналитики?



Веб-ресурсы:  
блоги, СМИ,  
социальные сети



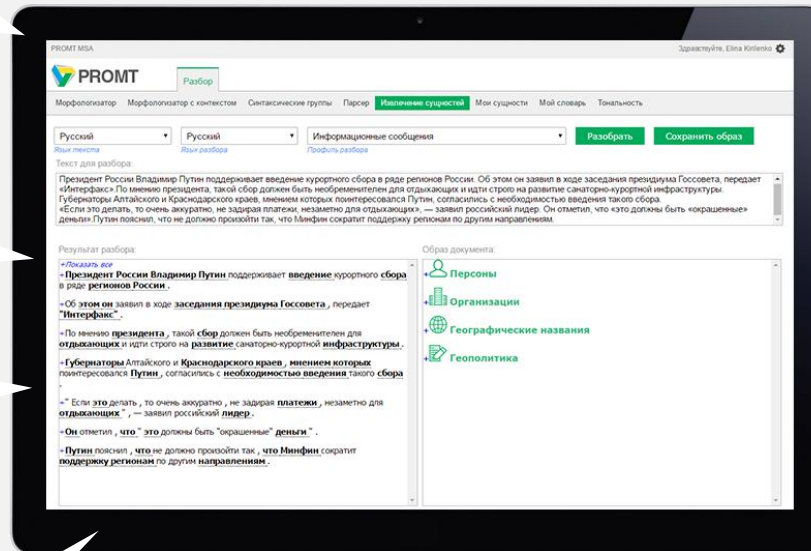
Архивы  
документов



Отчеты и  
измерения



Контент,  
сгенерированный  
пользователями



## АЛГОРИТМ АНАЛИЗА ПЕРЕВОДИМОЙ ИНФОРМАЦИИ:

Сбор информации →

Анализ информации  
на разных языках →

Извлечение  
сущностей →

Фильтрация, кластеризация,  
классификация,  
резюмирование  
информации



# Сложности анализа

При интеллектуальном анализе неструктурированных данных нужно учитывать:

- **Лингвистические технологии**
- **Морфологию**
- **Синтаксис и семантику языков**
- **Особенности текста (например, записи телефонных разговоров)**





# Лингвистический подход PROMT к анализу Big Data

- **Определение** анафорических **связей** ,  
установление **отношений** между извлеченными  
сущностями и фактами и их учет при анализе
- **Определение синонимов**
- **Выделение смысловой структуры документов**
- **Выделение сущностей** (в т.ч. именованных)  
и **фактов** (действий и высказываний)  
из текстов на иностранных языках
- **Определение тональности текста**,  
отдельных высказываний, тональной  
характеристики сущностей

# Функциональные возможности PROMT Analyzer Doc API

- извлечение фактографических данных из текстовых документов на естественных языках
- классификация документов
- интеллектуальный поиск

Выделяет ключевые объекты договора, сохраняет структурированную информацию в БД и может быть натренирован на выделение дополнительных объектов:



ТИП  
ДОГОВОРА



ДАТА И МЕСТО  
ЗАКЛЮЧЕНИЯ  
ДОГОВОРА



СТОИМОСТЬ  
ДОГОВОРА



ПРЕДМЕТ  
ДОГОВОРА



СРОК  
ДОГОВОРА



МЕСТО  
РАЗРЕШЕНИЯ  
СПОРОВ



СТОРОНЫ  
ДОГОВОРА

названия, представители с полномочиями, адреса/реквизиты и обязательства сторон по договору

# Функциональные возможности PROMT Analyzer News API

Позволяет собирать, индексировать и анализировать многоязычный контент из новостных источников

**СБОР:** важные новости с нужных ресурсов

**АНАЛИЗ:** выделение основных информационных объектов и автоматическое определение тональности

**ПОИСК:** База новостей с возможностью интеллектуального поиска по ней



ФАМИЛИИ И  
ИМЕНА  
ЛЮДЕЙ



ГЕОГРАФИЧЕСКИЕ  
НАЗВАНИЯ



ОРГАНИЗАЦИИ



АДРЕСА, E-MAIL,  
ТЕЛЕФОНЫ



ДАТЫ,  
ВРЕМЯ



ДЕНЕЖНЫЕ  
СУММЫ

# Какие задачи позволяет решать?

## Docs API



Юридическая экспертиза



Проверка на соответствие



Уменьшение рисков потери капиталов



Предупреждение мошенничества



Поддержка принятия решений



Предупреждение ошибок



Интеллектуальный поиск для выявления договоров с похожими условиями и/или совпадающими данными (контрагенты, суммы....)

## News API

Точный анализ информации, полученной от:



Сервисов для мониторинга сми и социальных сетей



Новостных агрегаторов и rss ридеров



Инструментов для сбора данных



Инструментов аналитики сми и социальных сетей

Повышение эффективности аналитиков, маркетологов, PR специалистов

# Преимущества



## СКОРОСТЬ

Высокая скорость работы,  
масштабируемость



## ОБУЧАЕМОСТЬ

На стороне заказчика



## ТОЧНОСТЬ

Точное выделение связей  
между объектами



## МНОГОЯЗЫЧНОСТЬ

Поддержка русского,  
английского и других языков



## ГИБКОСТЬ

Тонкая настройка под задачу заказчика



## БЕЗОПАСНОСТЬ

Работа в закрытом контуре

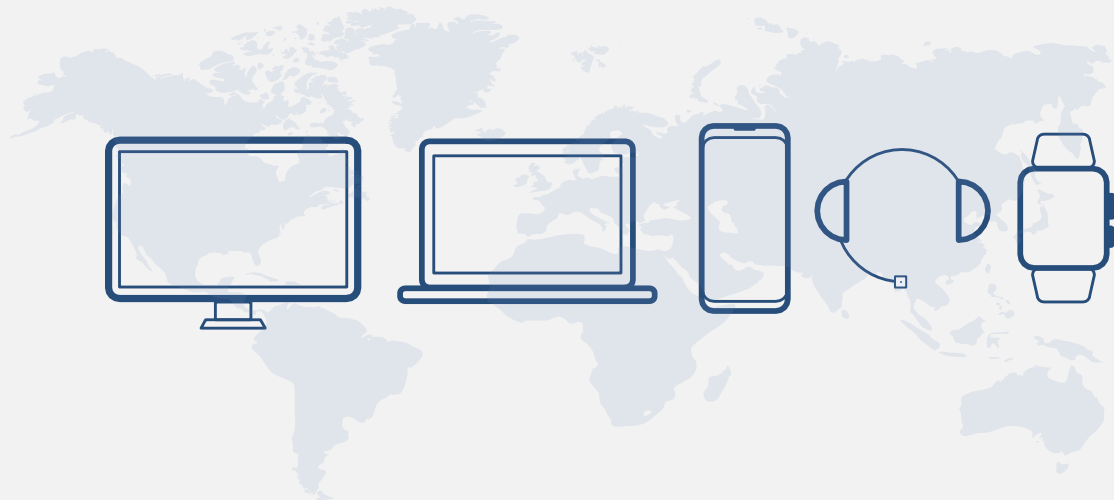
Призер соревнований по автоматическому извлечению информации  
из русскоязычных текстов **FactRuEval**

Финалист конкурса **Аналитического центра при правительстве РФ**

Специальная награда конкурса **«Лучшие информационно-аналитические инструменты»**



Технологии **PROMT** могут быть реализованы на разных ОС (**Windows, Linux, iOS, Android, Mac**) и могут быть интегрированы в сторонние программные комплексы благодаря **SDK**.



Компания поддерживает перевод для **25 языков** с использованием всех современных технологий перевода (аналитических, статистических, нейронных, гибридных). Также возможна разработка новых языковых пар (сроки от 3 до 6 месяцев).



# Бесплатный доступ к демо-стендам PROMT для всех участников конференции

Благодарим за внимание и ждем  
новых задач!

+7 (495) 580-48-48  
corporate@promt.ru

