TEXT MINING

интеллектуальный анализ архивных текстов для принятия эффективных решений

ООО «Институт оценки информации», г. Красноярск

Учредитель и генеральный директор – Павел Вадимович Полуян, к.ф.н., доцент СФУ

Полуян Павел Вадимович
генеральный директор ООО «ИОИ», канд. филос. наук, СФУ. Общее руководство и отраслевой маркетинг

Личаргин Дмитрий Викторович
канд. технических наук, СФУ, научный руководитель проекта, разработчик

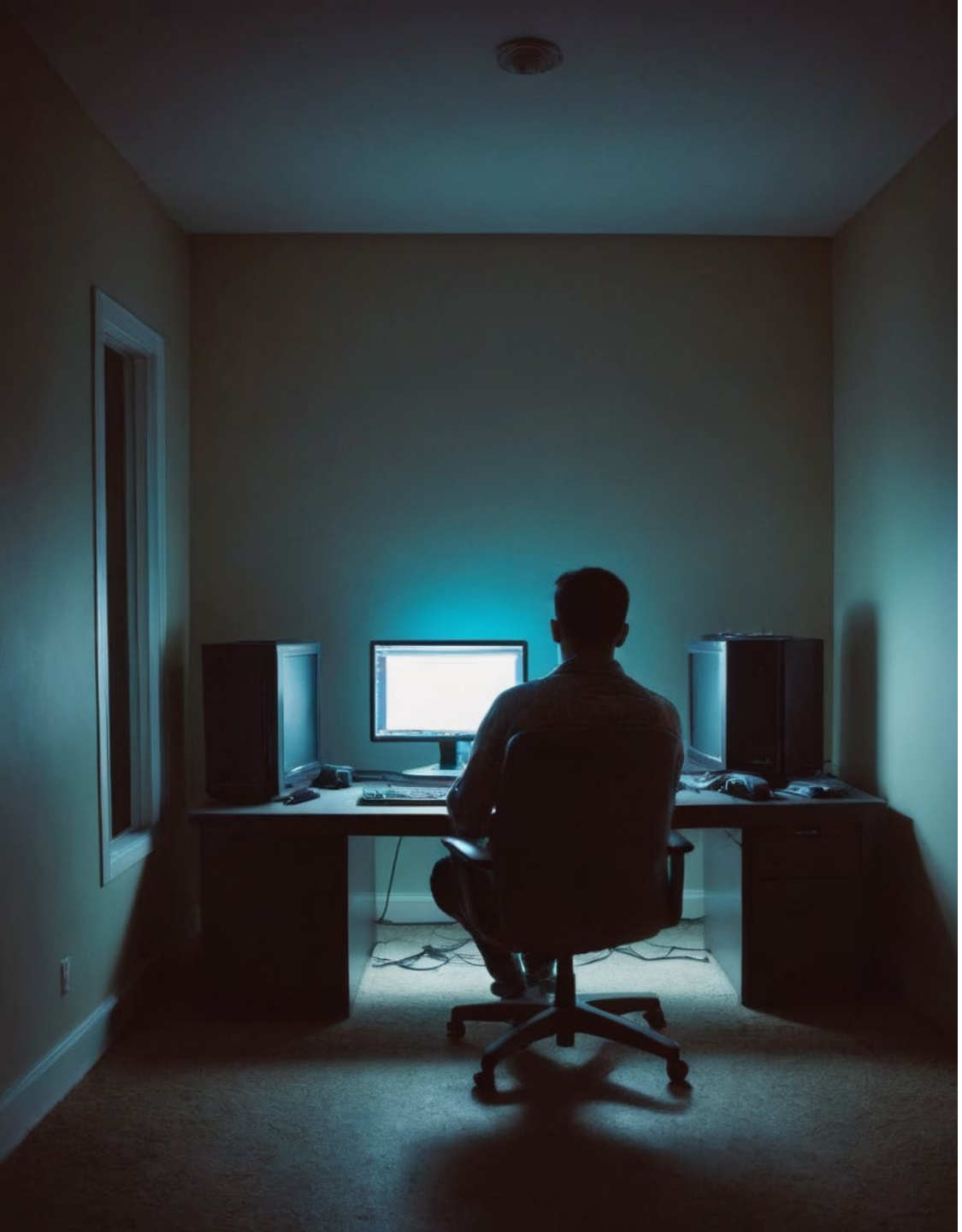
Шикунов Сергей Анатольевич
канд. физ-мат. наук, главный инженер, руководитель AI-направления

Корболеев Андрей Иванович
программист с 20 летним стажем

Бачурина Елена Петровна
концепт-менеджер, СФУ, специалист по продажам

Поротова Ольга Валерьевна
канд. экономических наук, СФУ, финансовый директор

Колодзинский Сергей Геннадьевич
сертифицированный оценщик
(«МСО» №0387 от 05.02.2008 г.)



Решаемая проблема

В корпорациях накопились обширные архивы оцифрованных неструктурированных файлов (текстов), где содержится текстовая информация, которая обрабатывается вручную.

Цели проекта

1 **Автоматически преобразовать** текстовую часть корпоративных архивов в базу данных - для интеграции, анализа и принятия эффективных решений.

2 Превратить архив в **объект интеллектуальной собственности** с последующей постановкой на баланс предприятия как нематериальный актив.

Описание продукта

Мы **решаем** эту **задачу** с помощью программного комплекса **«TEXT MINING»** на основе семантических технологий NLP и AI.

Разработан алгоритм TXT2DB, с его помощью создается цифровой двойник текста – это БАЗА ДАННЫХ. **Открывается возможность** универсального поиска по всем текстам архива.

Искусственный интеллект, обученный на текстах конкретных отчетов, работающий с архивом конфиденциальной информации, **способен отвечать** на поисковые запросы: Что? Где? Когда? Какая методика? Какие условия? Стоимость? Риски?

В перспективе *это будет* ***аналог ChatGPT****, который нацелен на работу с конкретным банком закрытых данных и обучается в процессе работы.*

Результаты интеллектуальной деятельности

Свидетельство о государственной регистрации
программы для ЭВМ Geosearch (Геологоразведка) модули TopViewUser, CrownsGeneratorAdmin
№ 2020616937 от 25.06.2020

Свидетельство о государственной регистрации
программы для ЭВМ Geosearch (Геологоразведка) модуль Web-реализация
№ 2021610071 от 12.01.2021

Свидетельство о государственной регистрации
программы для ЭВМ «Транспилятор TXT2DB» модуль Web-реализация,
№ 2023683861 от 03.11.2023

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2014615928

Генератор транскрипций

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2023685768

«Программный комплекс (ПК) «Транспилятор TXT2DB» модуль Web-реализация

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2013618592

Электронный словарь «Векторизованная семантическая классификация»

Резюме проекта

Проект	Интеллектуальный анализ архивных текстов для принятия эффективных решений
Категория	Информационные технологии
География	Россия, СНГ
Стадия	5 TRL Проверка основных технологических компонентов в реальных условиях 7 MRL Достигнуты первые продажи / пилотные внедрения (B2B)
Требуемые инвестиции	**10 млн. рублей**
Преимущества проекта	Нацеленность на большие тексты, невысокая стоимость продукта

Инвестиционное предложение

Объем инвестиций	**10**	**млн. рублей**
Доля в новой компании под проект	**15 %**	
Индекс прибыльности	**30 %**	
Срок фактической окупаемости для инвестора	**1 год**	

Взаимодействие с институтами развития

2022

получили от ФСИ грант на акселерацию: прошли учебу в московском Институте финансового развития бизнеса. В Тюмени участвовали в конкурсе INNOWEEK-02 – **победили и получили инвестиционный ресурс** от частного инвестора.

2023

успешно прошли акселератор INDUSTRIX, **заключили и выполнили пилотный проект** - коммерческий договор с ПАО «Газпромнефть» по разработке ПО с алгоритмом TXT2DB. Получили статус резидента ИЦ «Сколково». Вошли в список главных университетских стартапов.

2024

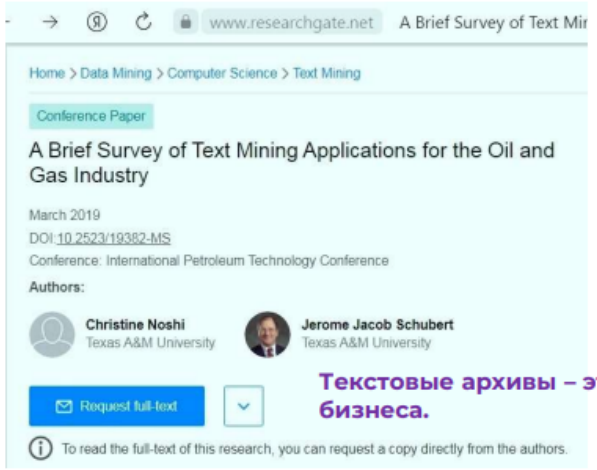
презентовали свой проект ООО «РН-КрасноярскНИПИнефть» и приняли участие в программе R&D Институт нефти и газа СФУ.

Анализ конкурентов

Компания	**ООО «ИОИ»** www.ieidp.ru	**ООО «ТриниДата»** www.trinidata.ru	**ООО «Мегапьютер Интеллидженс»** www.megaputer.ru
Продукт	ПК GEO-TEXT MINING	АрхиГраф	PolyAnalyst
Целевое использование	Ресурсный сектор, аналитика текстов ГРР отчетов и др. документов отрасли	Онтологические модели предметных областей	Предприятия торговли и фармацевтика
Превосходство	Нацеленность на большие тексты, создание баз данных на основе текстов и их интеграция Средний чек на продажу ПО – 1,5 млн.руб.	Разработанность продукта Лицензия приобретена в 2021 г. «Газпром Недра» за 1,8 млн.руб. – работа по готовым базам данных	Разработанность продукта Лидерство в области анализа отзывов, мнений, оценок и т.п. для целей маркетинга, поддерживается на 5 основных языках.
Недостатки	Продукт дорабатывается, с целью эксклюзивного охвата задач нефтегазовой отрасли НИОКР	Создание логистической витрины данных Высока цена	Аналитика большого количества малых текстов Высокая цена

Конкурентные преимущества

- Нацеленность на большие тексты
- Создание баз данных на основе текстов и их интеграция
- Невысокая стоимость продукта: средний чек на продажу ПО 1,5 млн руб.
- Высокая квалификация и опыт команды



Текстовые архивы – это информационные активы бизнеса.

Из США (2019)

TEXT MINING – это технология интеллектуальной обработки текстов, инструмент в системе принятия решений, нужный для анализа и сопоставлений.

Из Китая (2023)

Территория реализации проекта

Россия и страны СНГ, но проблема актуальна для всего мира и всех отраслей

Целевая аудитория

Клиент продукта – кто он?

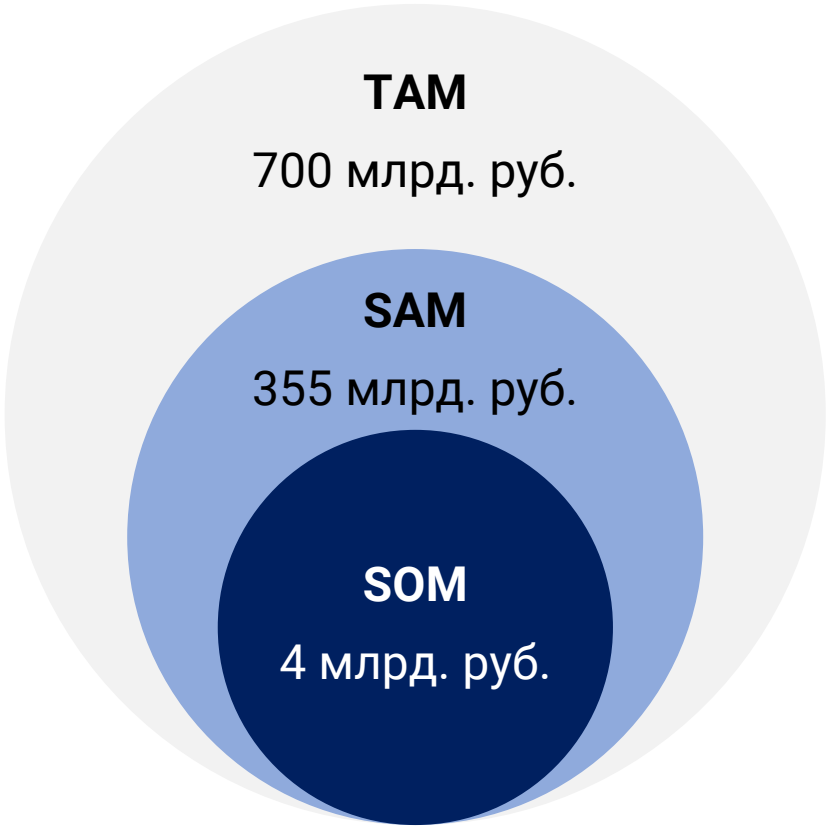
1. Ресурсная отрасль: недропользователи и геологоразведка
2. Корпорации с архивами документооборота
3. СМИ и медиахолдинги
4. Банки, инфраструктура финансового рынка
5. Научные учреждения
6. Связь и коммуникации
7. Поставщики услуг
8. Издательства и творческие агентства
9. Предприятия сферы культуры
10. Обладатели архивов (в т.ч. частные лица)

Экономические показатели

Параметр	2024 год	2025 год	2026 год
Интеллектуальная собственность	Доработка и **регистрация в РФ** ПК «Транспилятор TEXT2DB – GEOTEXT-MINING»	Доработка продукта до мировых стандартов	Доработка и **регистрация за рубежом** ПО «Транспилятор TEXT2DB – GEOTEXT-MINING»
Стадии развития продукта	Разработка и внедрение	Выход на рынок	Рост
Выручка	**3 000 000 руб.**	**7 000 000 руб.**	**10 000 000 руб.**
Количество сотрудников	10 человек	20 человек	50 человек

Ёмкость рынка

в РФ и за рубежом



Заказы и продажи

В 2025 г. начинается коммерциализация: продажа услуги предприятиям нефтегазовой отрасли – **три контракта в год окупают вложения**. Далее – мультипликация решений для предприятий других отраслей.

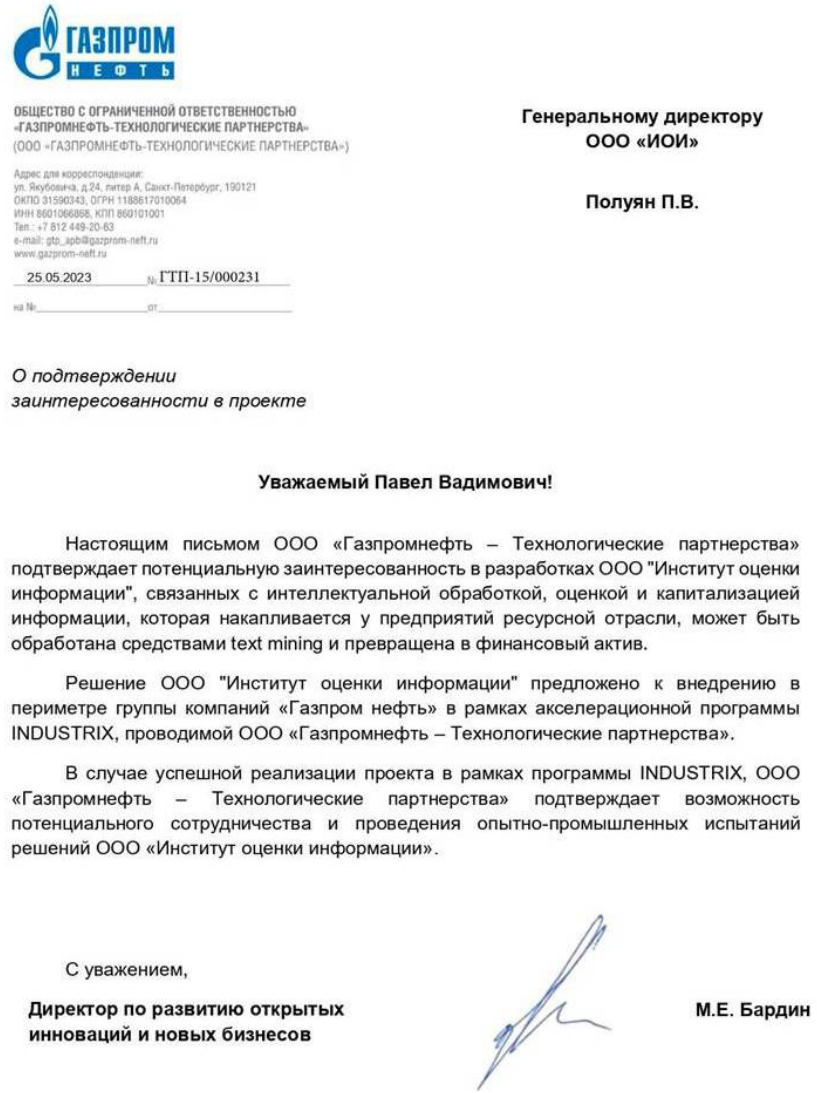
Услуга включает продажу лицензии на использование ПК, дообучение и настройку ИИ, послепродажное обслуживание и оценку ПК как нематериального актива заказчика.

Заказы и продажи

Текущая стадия продукта – TRL 5
Работоспособность технологии может быть продемонстрирована на детализированном макете в условиях, приближенным к реальным

Переход технологии на следующие уровни TRL 6 и 7 - по результатам пилотного проекта проводим доработку технологии до стадии промышленного образца

ГАЗПРОМ НЕФТЬ | НАУЧНО-ТЕХНИЧЕСКИЙ ЦЕНТР



ГАЗПРОМ НЕФТЬ

ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ
«ГАЗПРОМНЕФТЬ-ТЕХНОЛОГИЧЕСКИЕ ПАРТНЕРСТВА»
(ООО «ГАЗПРОМНЕФТЬ-ТЕХНОЛОГИЧЕСКИЕ ПАРТНЕРСТВА»)

Адрес для корреспонденции:
ул. Якубовича, д.24, литер А, Санкт-Петербург, 190121
ОКПО 31590343, ОГРН 1188617010064
ИНН 8601066868, КПП 860101001
Тел.: +7 812 449-20-63
e-mail: gto_apb@gazprom-neft.ru
www.gazprom-neft.ru

25.05.2023 № ГТП-15/000231
на № от

Генеральному директору
ООО «ИОИ»

Полуян П.В.

О подтверждении
заинтересованности в проекте

Уважаемый Павел Вадимович!

Настоящим письмом ООО «Газпромнефть – Технологические партнерства» подтверждает потенциальную заинтересованность в разработках ООО "Институт оценки информации", связанных с интеллектуальной обработкой, оценкой и капитализацией информации, которая накапливается у предприятий ресурсной отрасли, может быть обработана средствами text mining и превращена в финансовый актив.

Решение ООО "Институт оценки информации" предложено к внедрению в периметре группы компаний «Газпром нефть» в рамках акселерационной программы INDUSTRIX, проводимой ООО «Газпромнефть – Технологические партнерства».

В случае успешной реализации проекта в рамках программы INDUSTRIX, ООО «Газпромнефть – Технологические партнерства» подтверждает возможность потенциального сотрудничества и проведения опытно-промышленных испытаний решений ООО «Институт оценки информации».

С уважением,

Директор по развитию открытых инноваций и новых бизнесов **М.Е. Бардин**

ООО «Институт оценки информации»

интеллектуальная обработка и коммерческая оценка любых оцифрованных данных с целью их использования в системах принятия решений и постановки на баланс предприятия в качестве нематериальных активов.

+7 913 839 49 14

polyan2002@mail.ru

ieidp.ru