



EZEE PURCHASE

**Комплексное решение
по автоматизации закупочных процедур**

ЧТО УСПЕЛИ СДЕЛАТЬ?

1. Настраиваемый комплекс алгоритмов и ML моделей для поиска и скоринга поставщиков.
2. Фронт [React App](http://93.157.254.203:7778/)
<http://93.157.254.203:7778/>
3. Схемы процесса в Visio для упрощения встройки в существующие процессы Северстали
4. Прочее: юр.риски, CJM, NextSteps по развитию каждого модуля

КАК ЭТО РАБОТАЕТ?



ИЗ ЧЕГО СОСТОИТ РЕШЕНИЕ?

1. Алгоритм парсинга номенклатуры

Алгоритм, автоматически распарсивающий любую номенклатуру заданной структуры

2. Алгоритм поиска URL

Алгоритм, пополняющий справочник площадок для каждой номенклатуры

3. ML модель классификации URL

Алгоритм верификации и ранжирования площадок на основе метаданных (ML модель) и % охвата номенклатурных позиций

4. Алгоритм поиска поставщиков

Настраиваемый алгоритм поиска поставщиков и парсинга данных по ним.



ИЗ ЧЕГО СОСТОИТ РЕШЕНИЕ?

Jupyter Notebook

5. Процедура обогащения профиля поставщика доп. данными

Алгоритм поиска поставщика по наименованию / ИНН в открытых базах данных и обогащения доп. данными

6. ML - модель приоретизации поставщика

Алгоритм фильтрации поставщиков по существующим признакам (ликвидация, процедура банкротства) и ранжирование по скорингу ML модели

TBD

7. Удобный интерфейс для работы с базой поставщиков

GUI, позволяющий менеджеру по закупкам в несколько кликов получить всю необходимую информацию

PPT

8. Юридический риски и решения

1. АЛГОРИТМ ПАРСИНГА НОМЕНКЛАТУРЫ

Что делает модуль?

Распарсивает номенклатуру по признакам:

- Наименование
- Стандарт
- Характеристики

Настройка: не требуется

Развитие: при наличии номенклатуры другой структуры

Стек: Python: re, Pandas

На входе:

* Рукав с текстильным каркасом
B(II)-16-25-38-ХЛ ГОСТ 18698



На выходе:

Наименование	Стандарт	Характеристики
Рукав с текстильным каркасом	ГОСТ 18698	B(II)-16-25-38-ХЛ

2. АЛГОРИТМ ПОИСКА URL

Что делает модуль?

1. осуществляет поиск каждой номенклатурной позиции в поисковой системе
2. выбирает URL + метаданные данного URL + сохраняет HTML страницу
3. записывает в справочник порталов, релевантных для данной номенклатурной позиции

Зачем?

Экономия время на поиске и записи новых площадок с поставщиками

Стек: Python: re, Pandas

Настройка: Глубина поиска определяется автоматически – до достижения или по лимиту просмотра/ времени поиска.

Развитие: использование других поисковых систем, например, китайской 百度一下, 你就知道 (baidu.com)

На входе:

Наименование номенклатурной позиции

№		Наименование	Стандарт	Чистое наименование	Характеристики
0	1	* Манжета М50х70 ГОСТ 22704	ГОСТ 22704	Манжета	М50х70

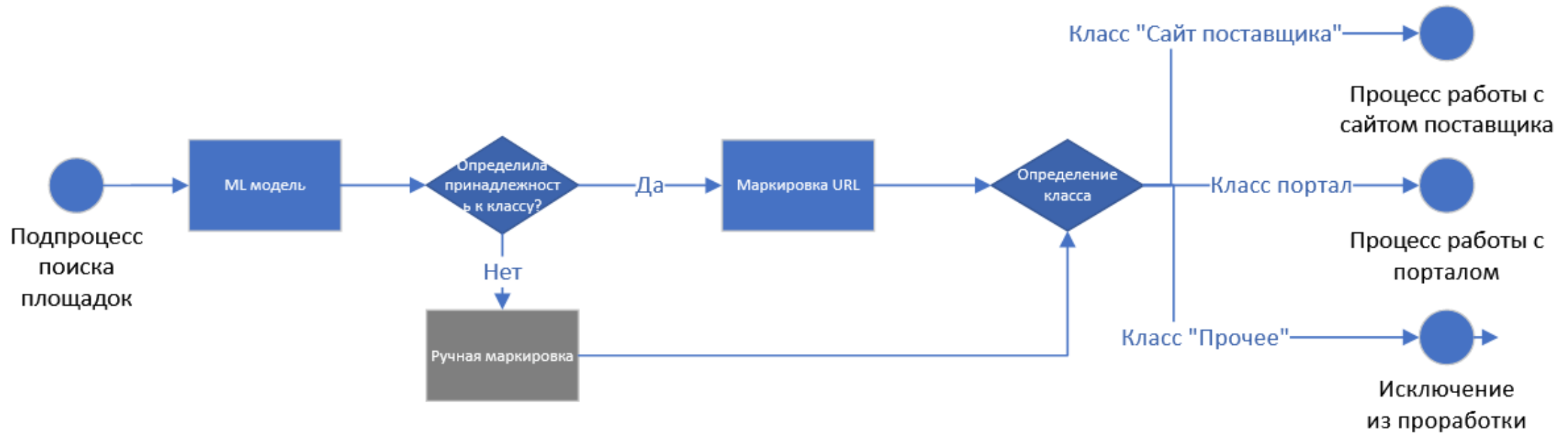


Наименование		URL	Описание
0	.	https://seal-is.com	Компания СИАЛ занимается производством РТИ (ре...
1	.	https://oniks-krep.ru	Продажа крепежа: болты, винты, заклепки, гайки...
2	.	https://uniform-met.ru	➤ Официальный сайт - «Юниформ Металл» ➤ На рын...
3	.	https://ms-74.ru	Автозапчасти для дорожно-строительной техники,...
4	.	https://www.pulscen.ru	Цены на товары и услуги в разделах: материалы,...

На выходе:

Наименование номенклатурной позиции –
URL вероятного поставщика или площадки -
метаданные - HTML

3. ML МОДЕЛЬ КЛАССИФИКАЦИИ URL: ПРОЦЕСС



Реализовано



Предлагается к реализации

3. ML МОДЕЛЬ КЛАССИФИКАЦИИ URL

Что делает модуль?

1. ML модель получает на вход URL площадок из выдачи вместе с метаданными по каждому URL + текст HTML
2. Метаданные проходят предобработку, лемматизацию, HTML преобразуется в текст
3. На основе заранее обученной модели , происходит классификация URL по 3 категориям: сайт поставщика, площадка, прочее

Зачем?

Экономим время на отсеве нерелевантных площадок

Стэк: html text classification using BERT transformers

Развитие: возможно сделать модель , которая будет непрерывно улучшаться на основе корректировки рез-та

На входе:

по конкретной позиции = «Гайка..». Индекс URL + текстовый признак URL

Индекс URL

	text
0	Изготовление деталей на автоматах продольного ...
1	ГайкиПродажа металлопроката оптом и в розницу ...
2	гайка шестигранная DIN934 M6 40 шткупить по вы...

Индекс URL

	text	target
0	Изготовление деталей на автоматах продольного ...	0
1	ГайкиПродажа металлопроката оптом и в розницу ...	0
2	гайка шестигранная DIN934 M6 40 шткупить по вы...	1

На выходе:

То же + соответствие классу (Портал)

3. ML МОДЕЛЬ КЛАССИФИКАЦИИ URL: ДЕТАЛИ

Что сделано:

1. Проанализировали существующие инструменты классификации аналогичных данных и выбрали `html text classification using BERT transformers`
2. Мы разметили 20+ уникальных URL из полученных в результате парсинга поисковой выдачи (прошлый этап) по 3 классам: URL сайта поставщика, URL портала (Желательно разметить 200+)
3. Провели обучение модели. Метрика качества – высокая, но необходимо увеличить объём обучающей выборки.

Epoch20 / 20

```
100%|██████████| 9/9 [00:00<00:00, 16.29it/s]
100%|██████████| 3/3 [00:00<00:00, 20.94it/s]
```

```
Training loss: 0.400
Test loss: 0.479
```

```
100%|██████████| 3/3 [00:00<00:00, 20.86it/s]
```

```
Test Accuracy: 0.800
```

4. АЛГОРИТМ ПОИСКА ПОСТАВЩИКОВ НА ПОРТАЛЕ

Что делает алгоритм?

1. Проверяет, есть ли площадка, которая выпала по результатам выдачи поисковой системы, в справочнике площадок с настроенным парсингом
2. Если да (сейчас алгоритм настроен под 1 площадку, под остальные можно адаптировать) то идет поиск поставщиков на этой площадке.
3. По каждому найденному поставщику заносится информация с портала.

Зачем?

Экономим время на поиске поставщиков на площадках и занесении информации в CRM

Стэк: Python: sklearn, nltk, pymorphy2

На входе: 1 номенклатурная позиция -

№		Наименование	Стандарт	Чистое наименование	Характеристики
0	1	* Манжета М50х70 ГОСТ 22704	ГОСТ 22704	Манжета	М50х70



		Товар	Поставщик	ИНН
0		* Манжета М50х70 ГОСТ 22704	ООО Стальметурал	[6670370745]
1		* Манжета М50х70 ГОСТ 22704	БСМ-Металл	[5410069100]
2		* Манжета М50х70 ГОСТ 22704	БСМ-Металл Москва	[5410069100]
3		* Манжета М50х70 ГОСТ 22704	«Промрезинотехника»	[2464257585]
4		* Манжета М50х70 ГОСТ 22704	Импульс	[5024131298]

На выходе: связка номенклатурная позиция — наименование и др. признаки поставщика

5. ПРОЦЕДУРА ОБОГАЩЕНИЯ ПРОФИЛЯ ПОСТАВЩИКА ДОП. ДАННЫМИ: СХЕМА



5.1 ПРОЦЕДУРА ОБОГАЩЕНИЯ ПРОФИЛЯ ПОСТАВЩИКА ДОП. ДАННЫМИ: ВНЕШНИЕ ИСТОЧНИКИ

Что делает модуль?

1. Выбирается список поставщиков, сформировавшихся после предыдущих этапов
2. По очереди ищет ИНН каждого поставщика
3. Обращается к базе ФНС по адресу <https://egrul.itsoft.ru> (Либо брать из ФНС напрямую – платно), обогащаем профиль - Статус (Работает/ Ликвидирован/ В процессе банкротства etc.), Уставный капитал , RUB Дебиторская задолженность, RUB Кредиторская задолженность, RUB
4. Обращаемся к DaData, обогащаем профиль (контакты + др. информация)

Зачем?

Экономим время на проверке актуальности поставщика
Готовим данные по поставщику для скоринга

Стэк: Python: sklearn, nltk, pymorphy2

На входе:

по конкретной позиции = «Гайка..». Индекс URL + текстовый признак URL

TBD



На выходе:

То же + соответствие классу (Портал)

5.2 ПРОЦЕДУРА ОБОГАЩЕНИЯ ПРОФИЛЯ ПОСТАВЩИКА: ML МОДЕЛЬ ОПРЕДЕЛЕНИЯ ПРЯМОГО ПОСТАВЩИКА

Что делает модуль?

1. Выбирается список поставщиков, сформировавшихся после этапа 4
2. Определяет принадлежность каждого поставщика к определенному классу

Что успели сделать?

1. Промаркировать 10+ URL прямой поставщик/посредник (очень мало, **нужно 200+**)
2. Обучили для примера модель. Ассурасу точно можно увеличить при увеличении обуч. выборки и подбора гиперпараметров

```
Epoch20 / 20
100%|██████████| 4/4 [00:00<00:00, 16.73it/s]
100%|██████████| 2/2 [00:00<00:00, 22.72it/s]

Training loss: 0.174
Test loss: 0.644

100%|██████████| 2/2 [00:00<00:00, 22.59it/s]

Test Accuracy: 0.667
```

На входе:

по каждому URL класса «поставщик» -
Индекс URL + текстовый признак URL

		text
0	заказать шплинт,шайбу,гровер О нас Новости Кон...	
1	Главная страница Торговый дом РТИ (Санкт-Петер...	
2	Рукав газовый (шланг) купить в Москве со скид...	
3	Оборудование щеточное, Навесное оборудование д...	
4	Крепеж STRONG КАТАЛОГ О НАС ДОСТАВКА И ОПЛАТА ...	

100%|██████████| 5/5 [00:00<00:00, 32.70it/s]

↓

		text	target
0	заказать шплинт,шайбу,гровер О нас Новости Кон...		1
1	Главная страница Торговый дом РТИ (Санкт-Петер...		0
2	Рукав газовый (шланг) купить в Москве со скид...		1
3	Оборудование щеточное, Навесное оборудование д...		1
4	Крепеж STRONG КАТАЛОГ О НАС ДОСТАВКА И ОПЛАТА ...		1

На выходе:

То же + определенный класс (Target –
прямой поставщик/ посредник).

6. ML - МОДЕЛЬ СКОРИНГА НАДЕЖНОСТИ ПОСТАВЩИКА

Что делает модуль?

Предварительно обученная модель по признакам из базы предсказывает вероятность неполной/несвоевременной поставки, низкого качества поставляемых товаров

NB: приведен ПРИМЕР модели, условно обученной на данных (из модуля 4). Для корректной работы модели нужны ретро-данные о работе поставщиков

Зачем?

Экономит время на проверке поставщика
Снижает риск неполной/несвоевременной поставки

Снижает риск низкого качества поставляемых товаров

Стек: Python: pandas, lightgbm








TBD

7. ФРОНТ

[React App](#)

<http://93.157.254.203:7778/>

Поиск поставщиков

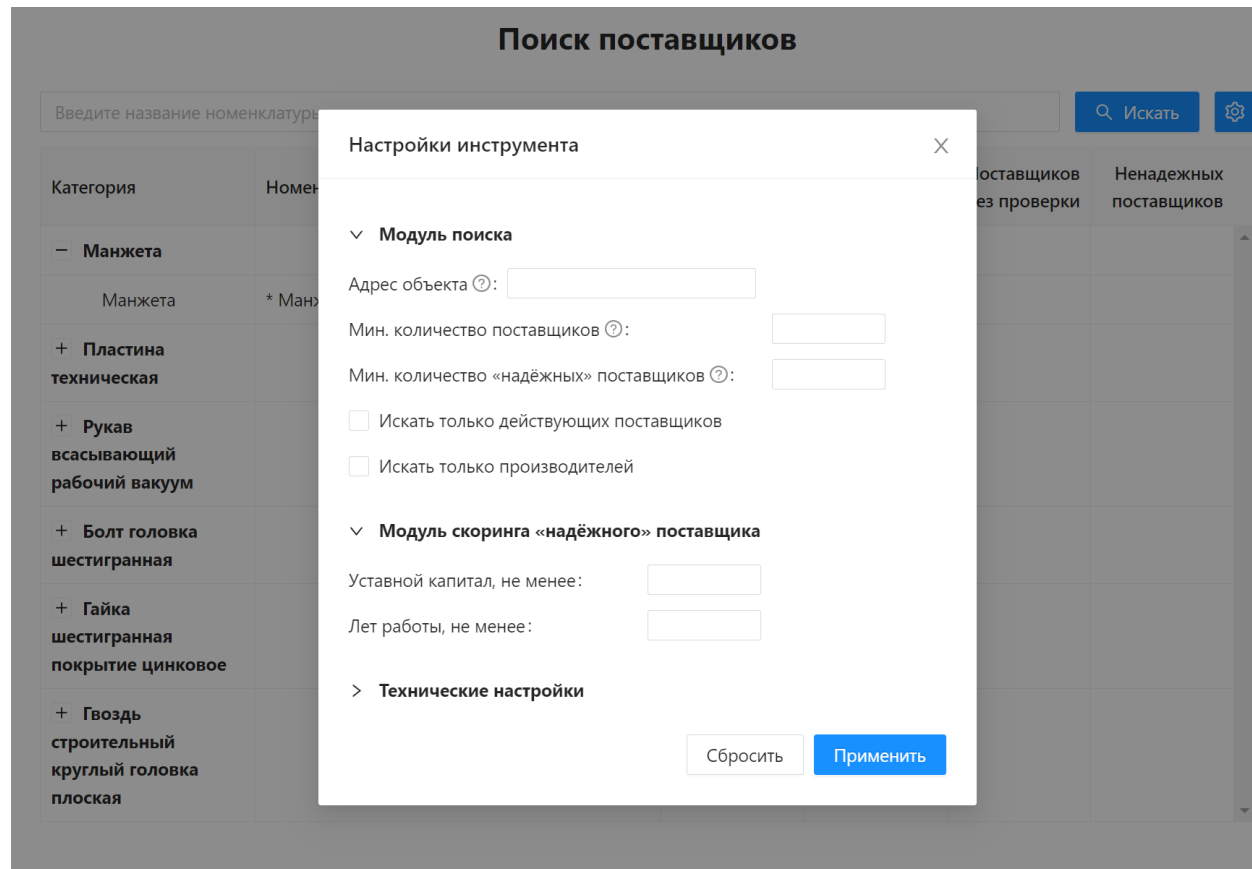
<input type="text" value="Введите название номенклатуры"/>						<input type="button" value="Искать"/>	
Категория	Номенклатура	Активных поставщиков	Надежных поставщиков	Поставщиков без проверки	Ненадежных поставщиков		
 Манжета							
Манжета	* Манжета M50x70 ГОСТ 22704						
 Пластина техническая							
 Рукав всасывающий рабочий вакуум							
 Болт головка шестигранная							
 Гайка шестигранная покрытие цинковое							
 Гвоздь строительный круглый головка плоская							

Стек: React, TypeScript

7. ФРОНТ

[React App](#)

<http://93.157.254.203:7778/>



Стек: React, TypeScript

8. ЮРИДИЧЕСКИЙ РИСКИ И РЕШЕНИЯ

Судебная практика:
Дело № А40-18827/2017

Нарушение исключительных прав
изготовителя базы данных

Основание: ст.1333, ст.1334, п. 3 ст. 1335.1 Гражданского кодекса РФ.

Решение: Получение разрешения правообладателя, однократное извлечение несущественной части базы данных.

Запрет извлекать и осуществлять
последующее использование
общедоступных персональных
данных

Основание: Статья 10.1. ФЗ № 152-ФЗ «О персональных данных» .

Решение: Получение согласия субъекта персональных данных . В случае нарушения - обезличить ПД.

Причинение вреда имуществу,
возмещение убытков
(Нагрузка на: сервер – порча, выход из
строя, на сайт: замедление работы)

Основание: ст. 15, ст. 1064 Гражданского кодекса РФ.

Решение: Использование ПО/алгоритма парсинга, минимизирующего нагрузку на сервер и сайт.

Неправомерный доступ к
компьютерной информации

Основание: ст. 272 Уголовного кодекса РФ

Решение: Избегать взлома паролей и охраняемой законом информации (коммерческая тайна, государственная тайна)

Команда EZEE

Александр Сергеев

Python, SQL
Product Owner,
DataAnalytic,
JunDS,
ProjectManagement
IPMA



Капитан, junDS

@Alexander_E_S

Алексей Шкиль

Python (flask, jinja,
pandas etc), c++, sql
JunDS
Project Management



Back

@alex_hlins

Никита Гетьман

Vue, Node, Project
Management,
проектирование и
дизайн веб-
интерфейсов,
JavaScript, HTML CSS,
Product Management



Front

@nikita_getman

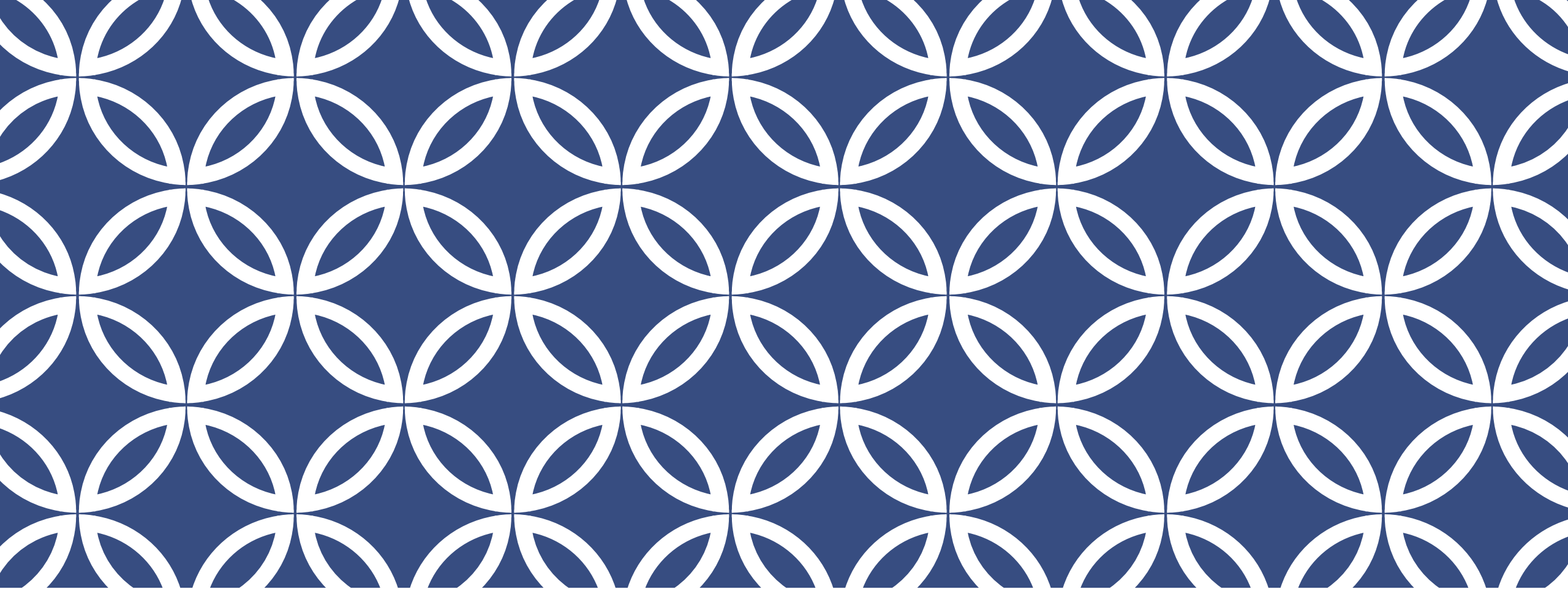
Дмитрий Меренков

SQL, Python, Civil,
Corporate Law,
Data Science



CJE

basel99@mail.ru
@BasBur



ПРИЛОЖЕНИЯ

Дополнительные слайды
Черновики
Идеи

NEXT STEPS

	Модуль	Что необходимо?	Какое улучшение?
1	Универсальный модуль парсинга номенклатуры	1. Полный список номенклатуры 2. Получить всю структуру уровней номенклатуры	1. Повысить универсальность алгоритма 2. Снизить количество запросов к внешним сайтам за счёт агрегации групп
2	Алгоритм поиска площадок	Получить от Заказчика уточненные критерии поиска площадок	Расширить периметр поиска и уточнить критерии поиска
3	ML модель ФИЛЬТРАЦИИ площадок	Получить от Заказчика список проверенных площадок как образец метаданных	1. На основе правильных площадок натренировать модель поиска релевантных площадок 2. Скорить метаданные по всем площадкам на релевантность
4	Алгоритм поиска поставщиков	1. Аналогично п.3. 2. stand-alone URL могут быть источником инф-и о поставщике?	1. Адаптировать алгоритм поиска поставщиков 2. Добавить ветку парсинга stand-alone URL 3. Предусмотреть полуавтоматический парсинг

NEXT STEPS

	Модуль	Что необходимо?	Какое улучшение?
5	Процедура обогащения профиля поставщика доп. данными	Какие ресурсы для проверки подключены у Заказчика? Есть ли API?	Дополнить информацию о поставщике новыми информативными признаками (бух.отчетность, наличие судебных дел, участие в гос.закупках и т.д.) Дополнить признаками о поставщике из реестров лицензий (открытые данные data.gov.ru)
6	ML модель	ретро-данные о работе поставщиков	Модель будет предсказывать неполную/несвоевременную поставку
7	Фронт	Получить от Заказчика клиентский путь и архитектуру использования решения	Разработка полноценного GUI (Front-end разработчик был занят на проекте до 8 марта)
8	Юридический риски и решения	Аналогично п.8	TBD

РИСКИ

	В чем риск?	Кого касается?	Как решаем?
1	Какие-то площадки не индексируются	Мелкие площадки/ сайты компаний?	Если поставщик не выкладывает свои услуги на портал, то он скорее всего не активен
2	Фронт-end выполняется браузером	Мелкие площадки/ сайты компаний?	RPA – алгоритм по ценным площадкам
3	Блокировка автоматизированных запросов		<ol style="list-style-type: none">1. Инструмент автоматизации работы с браузером Selenium2. Selenium может не подняться на сервере (РЕАЛИЗОВАН)!3. Регулировка частоты запросов
4			
5	Кто-то ЗАЯВЛЯЕТ что сделал универсальный		Донести до жюри что это невозможно. Сделать список сайтов которы

1.РАЗРАБОТАН УНИВЕРСАЛЬНЫЙ АЛГОРИТМ ПАРСИНГА НОМЕНКЛАТУРЫ

* Рукав с текстильным каркасом В(II)-16-25-38-ХЛ ГОСТ 18698



Рукав с текстильным
каркасом

Наименование

ГОСТ 18698

Стандарт

В(II)-16-25-38-ХЛ

Характеристики

ОПРЕДЕЛЯЕМ НАИМЕНОВАНИЕ

Анализируется фраза с начала: выбираются все слова, одно за другим (разделенные пробелом) и сверяем со справочником до тех пор, пока не встретится слово, которого нет в справочнике (1.2-120x150x12-1) или не встретится слово ГОСТ или СТАНДАРТ.

То, что соответствует п.1 вырезается из фразы и вставляется в колонку "Наименование"

Фраза: Рукав с текстильным каркасом В(II)-16-25-38-ХЛ ГОСТ 18698

База:

Рукав с текстильным
каркасом

Наименование

Стандарт

Характеристики

ОПРЕДЕЛЯЕМ СТАНДАРТ

1. Из фразы убирается слово "СТАНДАРТ"
2. Ищется "ГОСТ *" и/или "ИСО *" и или DIN*,
3. если найдено - забирается в ячейку и
4. вырезается из фразы.

Фраза: В(II)-16-25-38-ХЛ ГОСТ 18698

База:

Рукав с текстильным
каркасом

ГОСТ 18698

Наименование

Стандарт

Характеристики

ОПРЕДЕЛЯЕМ СТАНДАРТ

Оставшаяся фраза записывается в «Характеристики».

Характеристики также могут быть распарсены более детально, но для лота №3 это не требуется

Фраза: В(II)-16-25-38-ХЛ

База:

Рукав с текстильным
каркасом

ГОСТ 18698

В(II)-16-25-38-ХЛ

Наименование

Стандарт

Характеристики

Стэк: Python: re, Pandas

Развитие: при наличии номенклатуры другой структуры

Настройка: не требуется

ЭКРАН 1. НАСТРОЙКА СИСТЕМЫ

US: Закупщик хочет настроить, сколько и каких именно поставщиков необходимо искать:

Количество «надежных» поставщиков на 1 номенклатурную позицию: 20

Критерии «надежного» поставщика: уставный капитал, не менее: _____, лет работы, не менее: _____

Выберите адрес объекта поставки: Красноярск (индустрия 4.0.)

Максимальное расстояние до объекта поставки: любое

ОСНОВНОЙ ФУНКЦИОНАЛ СИСТЕМЫ

1. Посмотреть текущую наполненность по номенклатуре
2. Осуществит

ЭКРАН 2. ПОИСК

US:

1. Закупщик экспортирует список номенклатуры для поиска (обозначаем)
2. Закупщик вводит номенклатурную позицию в строку поиска

Обязательно:

Поле ввода

ЭКРАН 3. ВЫДАЧА

US

Закупщик смотрит на результаты, все параметры изучает, фильтрует как ему удобно, формирует итоговый список для обзвона или рассылки

Обязательно:

Доп. Функционал (Nice to have):

- показать поставщиков на карте
- построить маршрут (пусть пока будет автомобильный)

ЭКРАН 4 АНАЛИТИКА И ЛОГИСТИКА

US

1. Закупщик смотрит

- по каким территориям
- по каким номенклатурным группам
- сколько поставщиков

Поле ввода

ОТКРЫТЫЕ ВОПРОСЫ

1. сделать добавление инкрементами – прогнали всё – записали в базу.
Вводим новый запрос – база пополняется на результаты выдачи по данному запросу. Пока можно не делать логику перезаписи.

ДРАФТ СМ (ДЛЯ ПРОРАБОТКИ)

	ПОСТАНОВКА ЗАДАЧИ	ПОИСК ПОСТАВЩИКОВ		ОБСУЖДЕНИЕ	СОГЛАСОВАНИЕ
	Получение перечня требуемых товаров	Поиск информации	Проверка надежности	Финальные договоренности	Согласование поставщика
Actions		Ручной поиск информации в поисковой системе по каждой единицы товара из номенклатуры	Ручная проверка информации о статусе компаний, выбранных в качестве предполагаемого поставщика	Предполагается обсуждение долгосрочного сотрудничества, наличие корпоративных скидок, других необходимых условий с потенциальным поставщиком	
Process and channels	 <pre>graph LR; A[Email] --> B[Website]; B --> C[Website]; C --> D[Phone]; D --> E[Checklist];</pre>				
Problems		Временные затраты, необходимость проверять каждый маркетплейс	Временные затраты, необходимость проверять информацию на специализированных сайтах		Временные затраты, необходимость составления таблицы поставщиков для презентации вручную
Ideas / Opportunities		Предлагаем нами позволяет получать в автоматическом режиме в рамках одного запроса сразу множество маркетплейсов, на которых присутствует требуемый к поиску товар. Временные затраты существенно сокращаются.	Предлагаем нами алгоритм позволяет получать в автоматическом режиме информацию о юридических лицах из специализированных и открытых источников. Временные затраты существенно сокращаются.		Предлагаем нами алгоритм позволяет выгружать в автоматическом режиме базу данных. Временные затраты существенно сокращаются.

по п.1:

1. Универсальный модуль парсинга номенклатуры - оставляем как есть, т.к. не начем "обучаться" - нужны новые типы номенклатуры DONE!

2. Алгоритм поиска площадок - можем попробовать доработать. нужно найти скажем 30 реально промышленных торговых площадок, на которых закупают крупные конторы типа Северстали, скачать их метаданные. ЕСТЬ 1, ОСТАЛЬНЫЕ – КАСТОМИЗАЦИЯ (РУЧНИК)

3. ML модель фильтрации площадок - здесь можно прямо сейчас дорабатывать. Нужно на большем объёме данных научиться выделять правильные кластеры. Можно даже замутить обучение с подкреплением ☺. Типа когда менеджер по закупкам будет работать в системе и прокрывивать неправильно подобранные моделью площадки, модель будет дообучаться. Будет звучать как фиче. DONE, НО С ПОДКРЕПЛЕНИЕМ НАДО ПРЕДЛОЖИТЬ

4. Алгоритм поиска поставщиков - я бы сделал акцент относительно универсальности. Это слабая часть нашего решения. Пряма на мой взгляд киллер антифича) Можно же через индексацию сайта как-то поставщиков искать, это вроде универсальный способ. Или ещё как. DONE!

5. Процедура обогащения профиля поставщика доп. данными
Нужно найти максимум реестров и баз в открытом доступе по ЮЛ.

6. ML - модель приоретизации поставщика Можно попробовать найти в зарубежном интернете какой-нибудь кейс с прогнозированием работы поставщиков. Если найдем, это тоже будет киллер фича

7. Удобный интерфейс для работы с базой поставщиков
необходимо сейчас начинать делать. Фичей будет красивая аналитика (посмотрите статистику по поставщикам, смотрите - вот тут новые поставщики появились, не хотите ли встретиться?)
+ фичей будет возможность интегрировать наш визуал в их систему. Через какой-нибудь iframe или что-то такое. Смысл - что решение готово к использованию

8. Риски - ещё детальнее проработать риски, побольше практики + тренды западные

9. Можно еще фичи простенькие придумать - типа телеграмм бот который , который будет сигнализировать, если текущий поставщик начал банкротиться Умная аналитика.