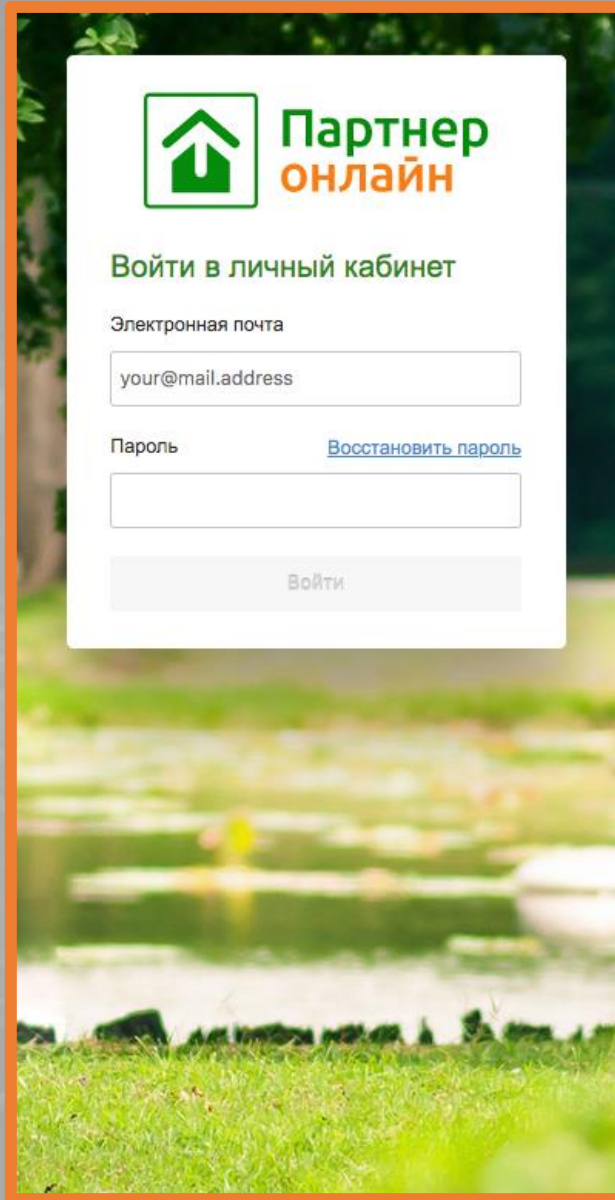


**VI Российский ипотечный конгресс**

# **СОКРАЩЕНИЕ ИЗДЕРЖЕК ЗА СЧЕТ РАСПОЗНАВАНИЯ РУКОПИСНЫХ ДОКУМЕНТОВ**

**Автор: Андрей Владыкин**  
проект “Посоветуй”

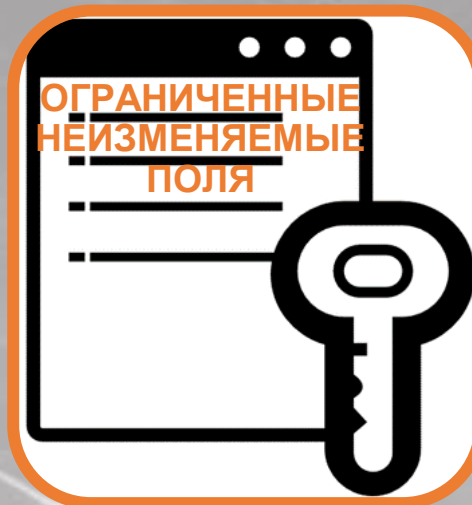


- Сбербанк:
  - **~2 000 человек** занимается распознаванием;
  - **2 минуты** тратит на верификацию каждого клиента;
  - **40 минут** тратил на распечатывание, сортировку, сканирование и ввод документов.
- ВТБ:
  - **~1 млрд документов** хранит в электронном виде.
- Газпромбанк:
  - **от 5 дней** для получения документов из архива

# ПОЧЕМУ ВСЁ ТАК МЕДЛЕННО РАЗВИВАЕТСЯ?



СТАНДАРТНЫЙ  
НАБОР  
ДОКУМЕНТОВ



ОГРАНИЧЕННЫЕ  
НЕИЗМЕНЯЕМЫЕ  
ПОЛЯ



РИСКИ СВЯЗАННЫЕ С  
ПЕРЕДАЧЕЙ ДАННЫХ НА  
РАССТОЯНИЕ

# Google

Open source решения







## Уже есть:

- электронные банковские архивы
- электронный документооборот
- кредитные конвейеры
- онлайн заявки
- мобильные приложения

## В ожидании:

- не оцифрованные архивы (вне указаний ЦБ)
- документооборот и кредитование юрлиц
- фото объектов залога с геопозицией

## Следовательно:

- рост финтех-рынка на 12% до 54 млрд рублей (Deloitte)

# КАК РАБОТАЕТ СИСТЕМА РАСПОЗНАВАНИЯ ДОКУМЕНТОВ

Скриншот интерфейса системы распознавания документов. Вверху — заголовок «Система распознавания документов». Ниже — таблица с данными о документе, включая поля для ввода и выбора. Внизу — график с координатами точек (X, Y) и легендой «IMG points: 51 TMP points: 58».

Особенности рукописного текста:

кыргызка 0.96 LL=0 — нет слова в русском языке

Кыргызка

сараяк 0.64 LL=1 — выбор в сторону буквы

САРАЯК

душанбе 0.42 LL=1 — выбор в сторону встречаемости

ДУШАНБЕ

1. находит **реперные точки** из шаблона в скане документа
2. кадрирует и **поворачивает** документ, до совпадения точек шаблона и скана
3. **находит изображения на скане** в местах, указанных на шаблоне
4. выделяет каждый “новый” элемент и **сравнивает** его с каталогом
5. **сверяет** результат поиска с библиотекой слов
6. слово с самым высоким рейтингом становится **результатом поиска**

## Возможности систем распознавания:

- распознавание **рукописного** текста в блоках, строках и ячейках,
- распознавание **ручных отметок** (чек-боксы и другие),
- распознавание и **сортировка рассыпанных документов** (распознает тип документа, раскладывает по-порядку),
- проверка **комплектации документов** на соответствие пакету-перечню
- **поиск** данных (ФИО, адреса, суммы) и сверка между собой,
- **создание отчетов** в стандартный форматах (excel, word, csv),
- распознавание **QR кодов** на документах.

Точность распознавания отметок, кодов и печатного текста - **99%**

**Страница 1 (начало анкеты)**

1. Дата рождения (ДД.ММ.ГГГГ): 

18	12	1973
----	----	------

 г.

2. Пол: ☐ мужской ☒ женский

3. Место Вашего рождения:

страна: 

Р	Ф				
---	---	--	--	--	--

область или республика: 

У	Д	М	У	Р	Т	С	К	А	Я
Р	Е	С	П	У	Б	Л	И	К	А

город или населенный пункт: 

И	Ж	Е	В	С	К				
---	---	---	---	---	---	--	--	--	--

4. Национальность (как Вы сами ее определяете): 

Р	У	С	С	К	А	Я						
---	---	---	---	---	---	---	--	--	--	--	--	--

5. Нынешнее гражданство: ☒ Россия

другая страна: 

--	--	--	--	--	--	--	--	--	--	--	--	--

6. Фактическое место Вашего сегодняшнего проживания:

область или республика: 


город или населенный пункт: 

М	О	С	К	В	А				
---	---	---	---	---	---	--	--	--	--

## Страница 1 (начало анкеты)

1. Дата рождения (ДД.ММ.ГГГГ):

2. Пол:

3. Место Вашего рождения:

страна:

область или республика:

город или населенный пункт:

4. Национальность (как Вы сами ее определяете):

5. Нынешнее гражданство:

другая страна:

6. Фактическое место Вашего сегодняшнего проживания:

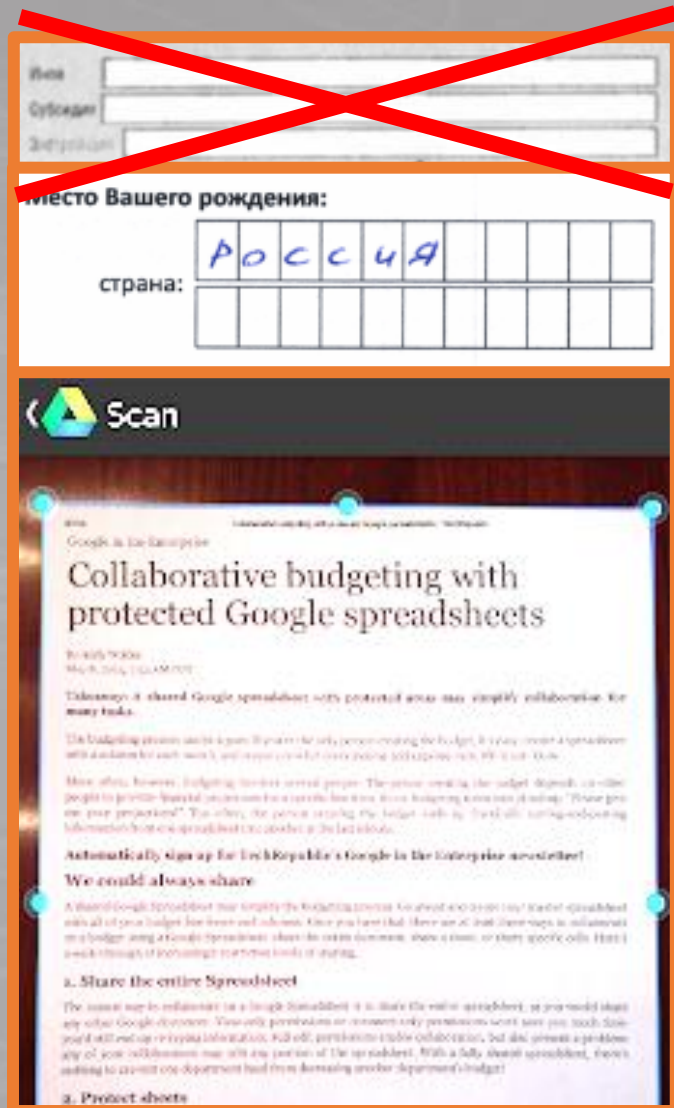
область или республика:

город или населенный пункт:

Изображение	Текст	Вероятность	Ближайшее
	18 12 1973	78.0 %	
	рф	99.0 %	
		100.0 %	
	удмуртская	37.0 %	удмуртская
	республика	45.0 %	республика
	ижевск	88.0 %	Ижевск
		100.0 %	
	русская	98.0 %	русская
		100.0 %	
		100.0 %	
		100.0 %	
	москва	70.0 %	москва
		100.0 %	
		100.0 %	
		100.0 %	
	4	32.0 %	

удмуртская
республика
ижмевск
русская
москва





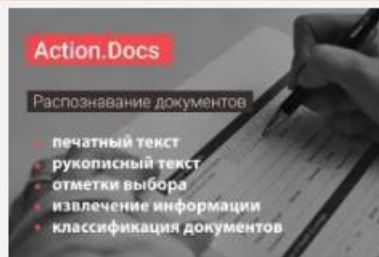
## Рекомендации:

- **большие** ячейки для рукописного текста
- **одна** клетка для одной буквы
- **синяя** ручка (не черная)
- *меньше печатей* на тексте

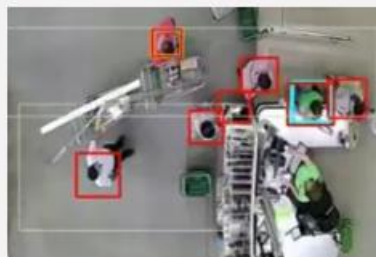
## Вот бы донести информацию:

(до риелторов и клиентов)

- качество анкет и сканов прямо влияет на скорость решения
- используйте специальные приложения (Google Disk)
- используйте свойства камер телефонов (кадрирование)
- больше света, меньше пальцев )))



ACTION.DOCS – РАСПОЗНАВАНИЕ  
РУКОПИСНЫХ ДОКУМЕНТОВ



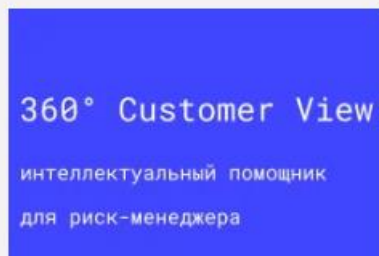
ТЕПЛОВЫЕ КАРТЫ – РАСПОЗНАВАНИЕ  
И ПОДСЧЁТ ПОСЕТИТЕЛЕЙ



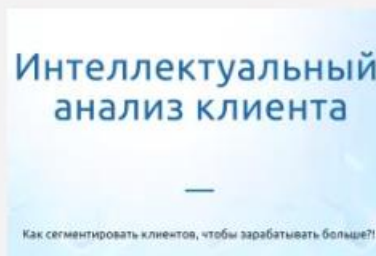
КРАУДСОРСИНГ – ОБРАБОТКА ИДЕЙ  
СОТРУДНИКОВ



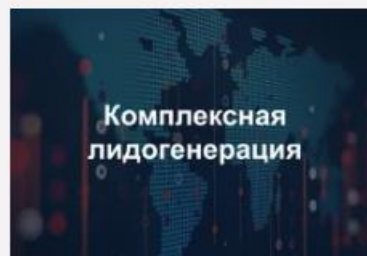
ЗАКАЗНАЯ РАЗРАБОТКА ПО



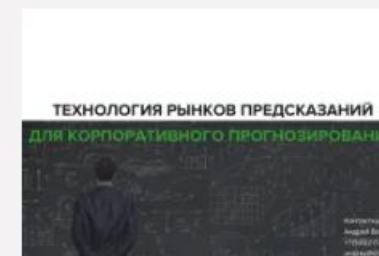
РИСК-АНАЛИЗ КОНТРАГЕНТА



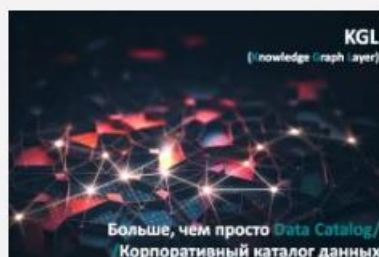
ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ  
КЛИЕНТА



КОМПЛЕКСНАЯ ЛИДОГЕНЕРАЦИЯ



КРАУДФОРКАСТИНГ – СИСТЕМА  
ПРЕДСКАЗАНИЙ



КОРПОРАТИВНЫЙ КАТАЛОГ ДАННЫХ



КОРПОРАТИВНЫЙ МАГАЗИН  
БРЕНДИРОВАННЫХ ТОВАРОВ



WIFI-РАДАР



МАРКЕТПЛЕЙС – САЙТ ДЛЯ  
РИЕЛТОРОВ