

Опыт развития ИИ решений для работы с документами ДОМ РФ

Орешкин Сергей

Директор департамента данных и новых технологий

DOM.IDP – сервис автоматизации работы с документами



Убираем ручной труд с заполнением всех полей для ипотечной заявки

Раньше

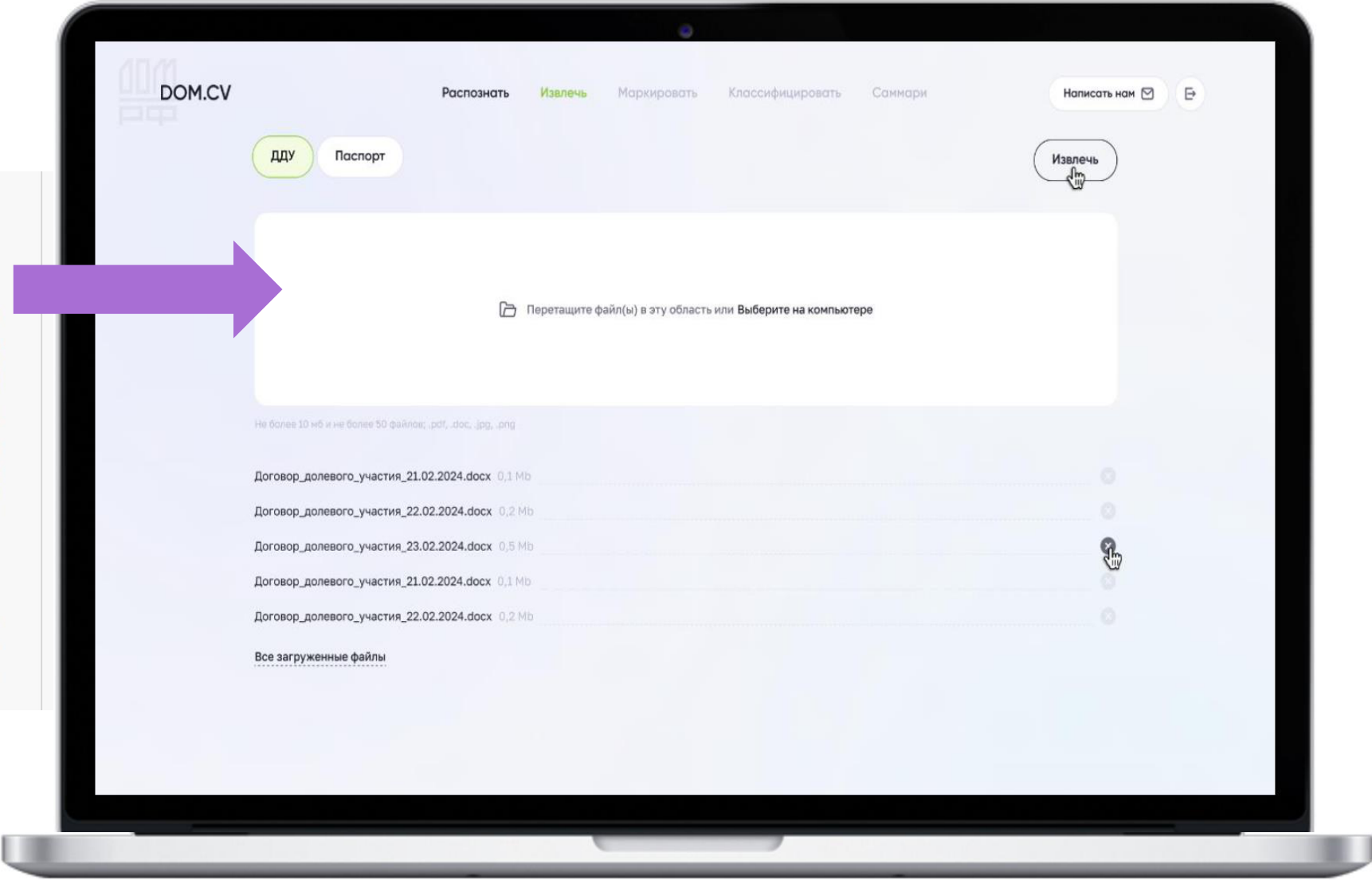
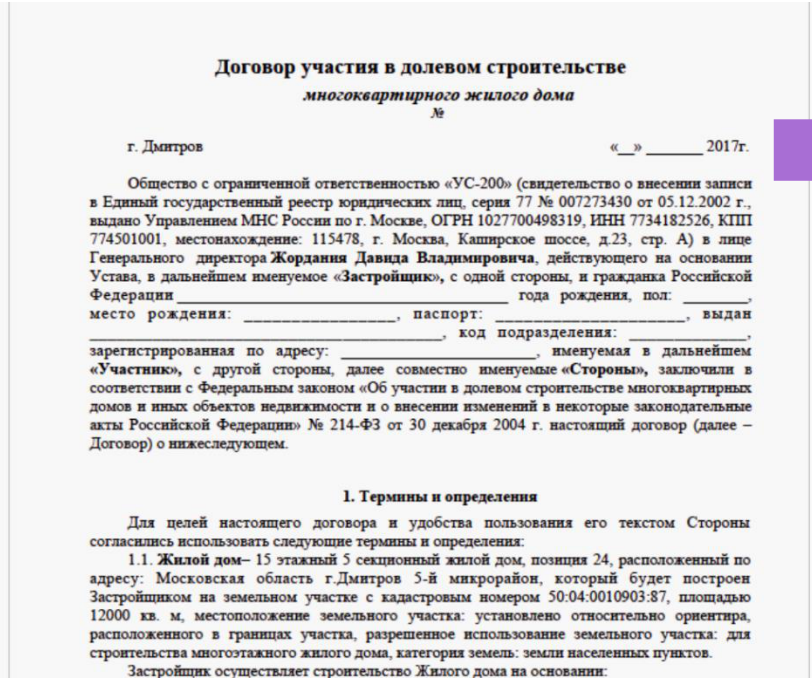
ручной ввод данных в ипотечный договор: ФИО, ИНН, стоимость объекта недвижимости и пр.

С DOM.IDP

«руками» загружаете только документ в систему и получаете уже заполненный финальный документ

Загрузите договор ДДУ или паспорт в интерфейс DOM.IDP

Система начнет распознавание содержимого документов



Проверьте распознанные системой поля

Значения автоматом подставляются в нужное поле



Бизнес требования к функционалу платформы DOM CV
(ДДУ, ПАСПОРТА)

DOM CV - Набор сервисов, которые помогают сотрудникам Банка работать с документами, а именно –

1. Распознавание (OCR) – это процесс преобразования изображения текста в машиночитаемый текстовый формат

2. Извлечение – поиск и извлечение из текста документа необходимых атрибутов (ФИО клиента, адреса, реквизиты и т.п)

3. Маркировка – выделение цветом в тексте документа важные слова и словосочетания для облегчения их поиска

4. Классификация – разделение документов по типам(акты, счета, договора и т.п) с присвоением каждому документу соответствующего признака.

5. Саммаризация – составление краткого содержания текста документа

В данном документе описаны бизнес требования к таким функциональным блокам сервиса DOM CV как –

• «Распознавание»

• «Извлечение» (пока только для ДДУ и Паспортов РФ)

А также описаны требования к визуальному отображению очереди обработки документов и стартовой страницы (рабочего стола) пользователя сервиса DOM CV.

Данный документ описывает требования к MVP продукта. Требования для рабочей версии могут изменяться.

1. Стартовая страница (рабочий стол)

1.1. Внешний вид стартовой страницы предназначен для выбора пользователем необходимого функционала и выглядит следующим образом

1. Распознавание (OCR) – это процесс преобразования изображения текста в машиночитаемый текстовый формат

2. Извлечение – поиск и извлечение из текста документа необходимых атрибутов (ФИО клиента, адреса, реквизиты и т.п)

3. Маркировка – выделение цветом в тексте документа важные слова и словосочетания для облегчения их поиска

4. Классификация – разделение документов по типам(акты, счета, договора и т.п) с присвоением каждому документу соответствующего признака.

5. Саммаризация – составление краткого содержания текста документа

В данном документе описаны бизнес требования к таким функциональным блокам сервиса DOM CV как –

• «Распознавание»

• «Извлечение» (пока только для ДДУ и Паспортов РФ)

А также описаны требования к визуальному отображению очереди обработки документов и стартовой страницы (рабочего стола) пользователя сервиса DOM CV.

Данный документ описывает требования к MVP продукта. Требования для рабочей версии могут изменяться.

1. Стартовая страница (рабочий стол)

1.1. Внешний вид стартовой страницы предназначен для выбора пользователем необходимого функционала и выглядит следующим образом

РаспознатьИзвлечьМаркироватьКлассифицировать

ДДУПаспортДоговор_долевого_участия_21.0

Назад

Объект

Помещение

1. ИНН застройщика24 0000 1234 65

2. Кадастровый номер50:15:0000000:153324

3. ФИО клиентаЧелбас Рыбак Иванович

4. ИНН клиента24 0000 1234 65

Клиент

Условия

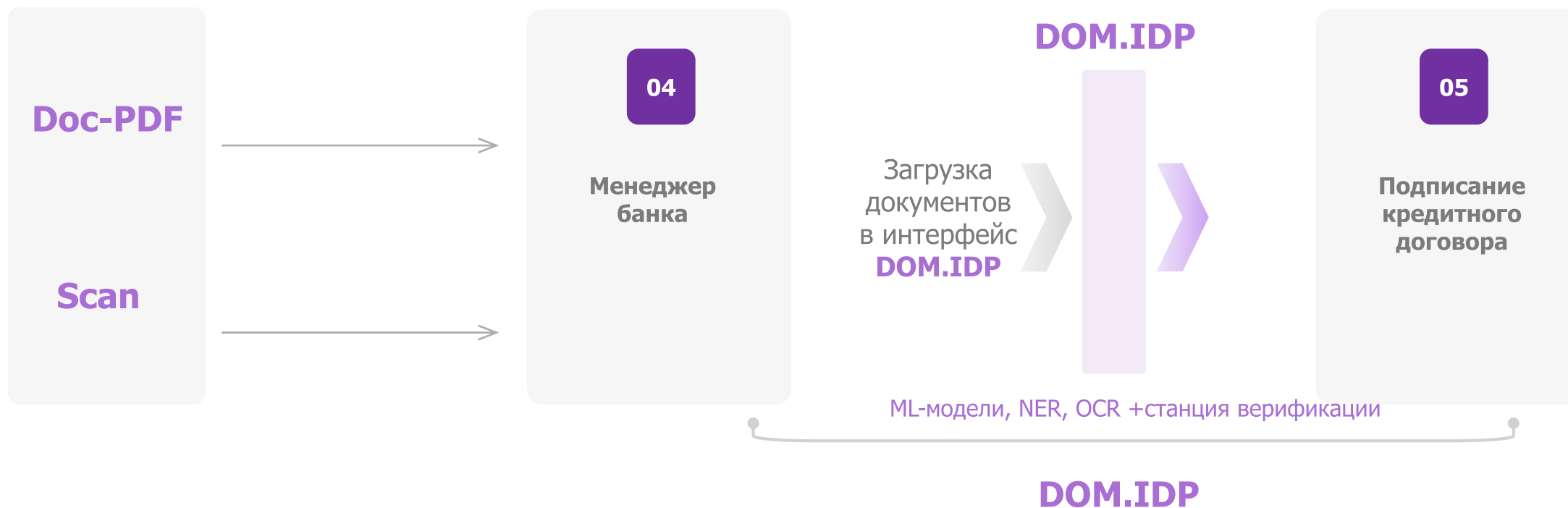
24 0000 1234 65

Финальный документ готов!

Автоматически создана: заявка на ипотеку, ипотечный договор,
форма одобрения объекта недвижимости



Как обрабатывает DOM.IDP на ДДУ



Что такое **intelligent document processing (IDP)*?**



Основной функционал IDP систем -



Распознавание (OCR) – превратим скан документа в редактируемый формат



Извлечение – найдем и вытащим нужные атрибуты из текста документа (например, ФИО, ИНН, адрес и т.д.)



Классификация – разложим документы по типам



Верификация – сотрудник проверит извлеченные данные



Маркировка – подсветим в тексте документа важные слова и словосочетания



Саммаризация – сформируем краткое содержание текста документа

CV. OCR (optical character recognition)



Скан ДДУ

ДОГОВОР № 13-...
участия в долевом строительстве

г. Москва «09» марта 2021 года

Общество с ограниченной ответственностью «Велестрой», записанное в Едином государственном реестре недвижимости (ЕГРН) МФННС № 46 по г. Москве 04.04.2008 года, ОГРН 1087746466950, ИНН 7709787790, КПП 7709787790, с местом нахождения по адресу: 125047, г. Москва, улица 2-ая Тверская-Ямская, дом 10, именуемое в дальнейшем «Застройщик», в лице генерального директора Златко Пенча, действующего на основании Устава, с одной стороны, и

Гр. РФ

именуемый в дальнейшем «Участник долевого строительства», с другой стороны, при совместном упоминании именуемые в дальнейшем «Стороны», руководствуясь Гражданским Кодексом Российской Федерации, Федеральным законом от 30.12.2004 г. № 214-ФЗ «Об участии в долевом строительстве многоквартирных домов и иных объектов недвижимости и о внесении изменений в некоторые законодательные акты Российской Федерации» (далее по тексту – Закон 214-ФЗ), заключили настоящий договор участия в долевом строительстве (далее – Договор) о нижеследующем:

1. ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

Применяемые в настоящем Договоре термины и определения имеют следующее значение:

1.1. **Многоквартирный дом** – многоквартирный жилой дом по адресу: г. Москва, ЗАО, район Фили-Давыдовское, улица Касташевская, вл.68, создаваемый Застройщиком, в том числе с привлечением денежных средств Участника долевого строительства, по Земельному участку с кадастровым номером 77:07:0005003:87, общей площадью 12 000 кв.м., категория земель: земли населенных пунктов, и на Земельном участке с кадастровым номером 77:07:0005003:50, общей площадью 234 кв.м., категория земель: земли населенных пунктов.

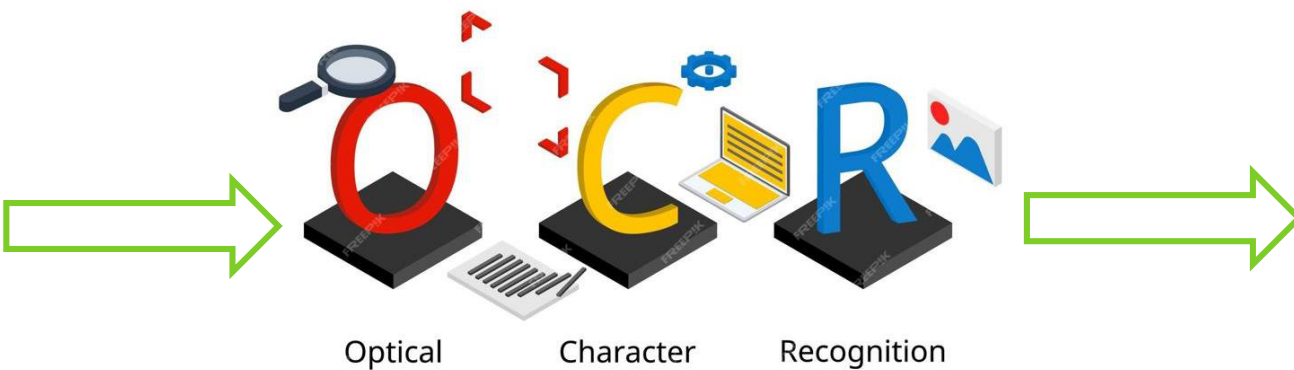
Основные характеристики многоквартирного дома указаны в Приложении № 3 к настоящему Договору.

1.2. **Объект долевого строительства (Квартира)** – изолированное, предназначенное для проживания граждан жилое помещение (квартира), имеющее индивидуальные характеристики, а также общее имущество в Многоквартирном доме, подлежащее передаче в собственность Участнику долевого строительства после получения Застройщиком разрешения на ввод в эксплуатацию Многоквартирного дома и входящее в состав Многоквартирного дома.

1.3. **Проектная (расчетная) площадь Объекта долевого строительства** – сумма площадей всех частей помещения, включая площадь помещений вспомогательного использования, а также балконов и лоджий (с применением понижающих коэффициентов: для лоджий – 0,5; для балконов – 0,3), которая определяется в соответствии с проектной документацией на Многоквартирный дом на дату подписания настоящего Договора.

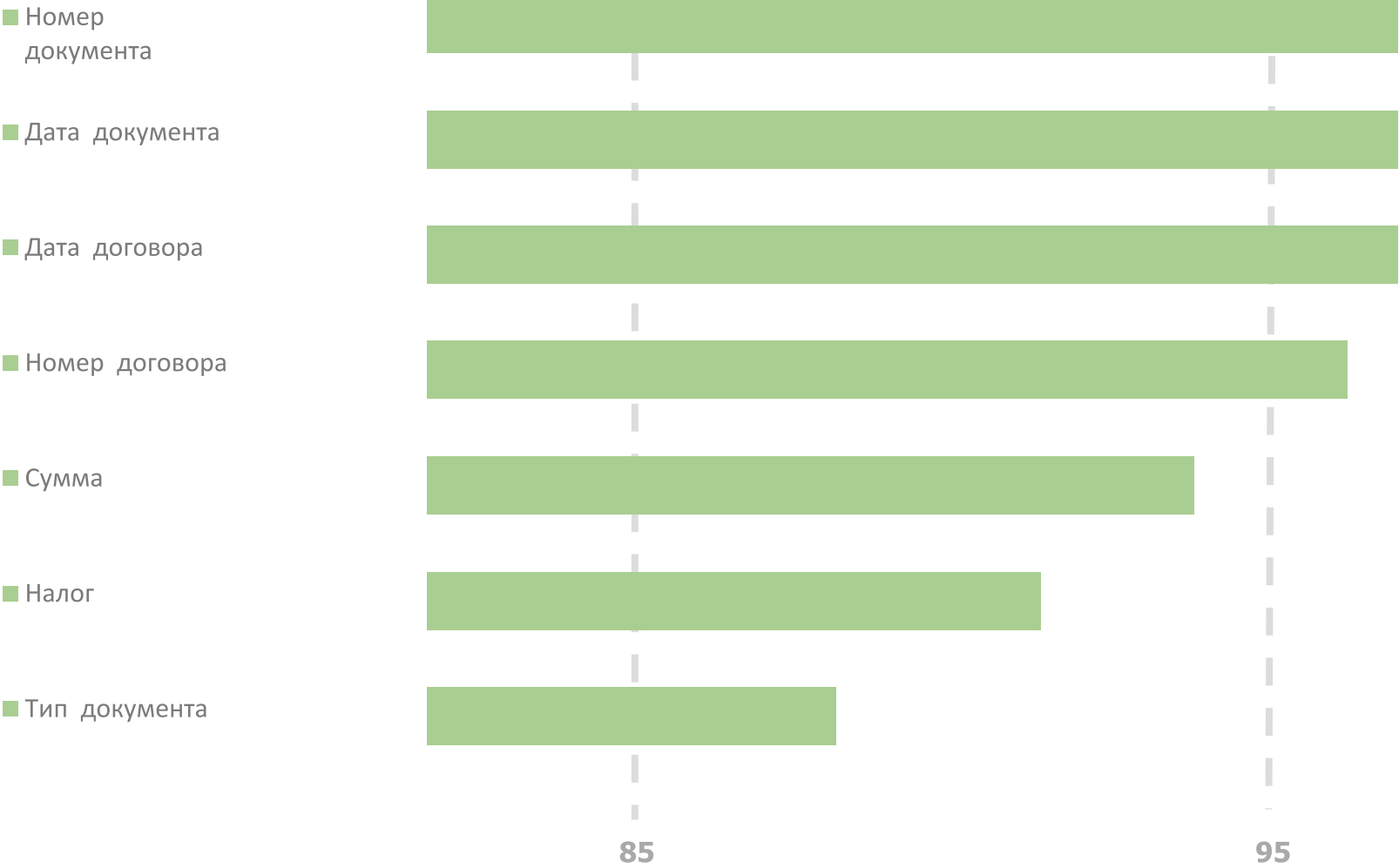
1.4. **Фактическая площадь Объекта долевого строительства** – сумма площадей всех частей помещения, включая площадь помещений вспомогательного использования, а также балконов и лоджий (с применением понижающих коэффициентов: для лоджий – 0,5; для балконов – 0,3), которая определяется по окончанию строительства Многоквартирного дома на основании проведенных Кадастровым инженером обмеров Объекта долевого строительства проводится в степени готовности согласно проектной документации и составленного технического плана на Многоквартирный дом.

1.5. **Кадастровый инженер** – лицо, уполномоченное в соответствии с действующим законодательством на выполнение кадастровых работ в отношении объектов недвижимого имущества, привлеченное Застройщиком к выполнению кадастровых работ в отношении Многоквартирного дома.



Tesseract

Точность извлечения атрибутов из бухгалтерских документов*, %

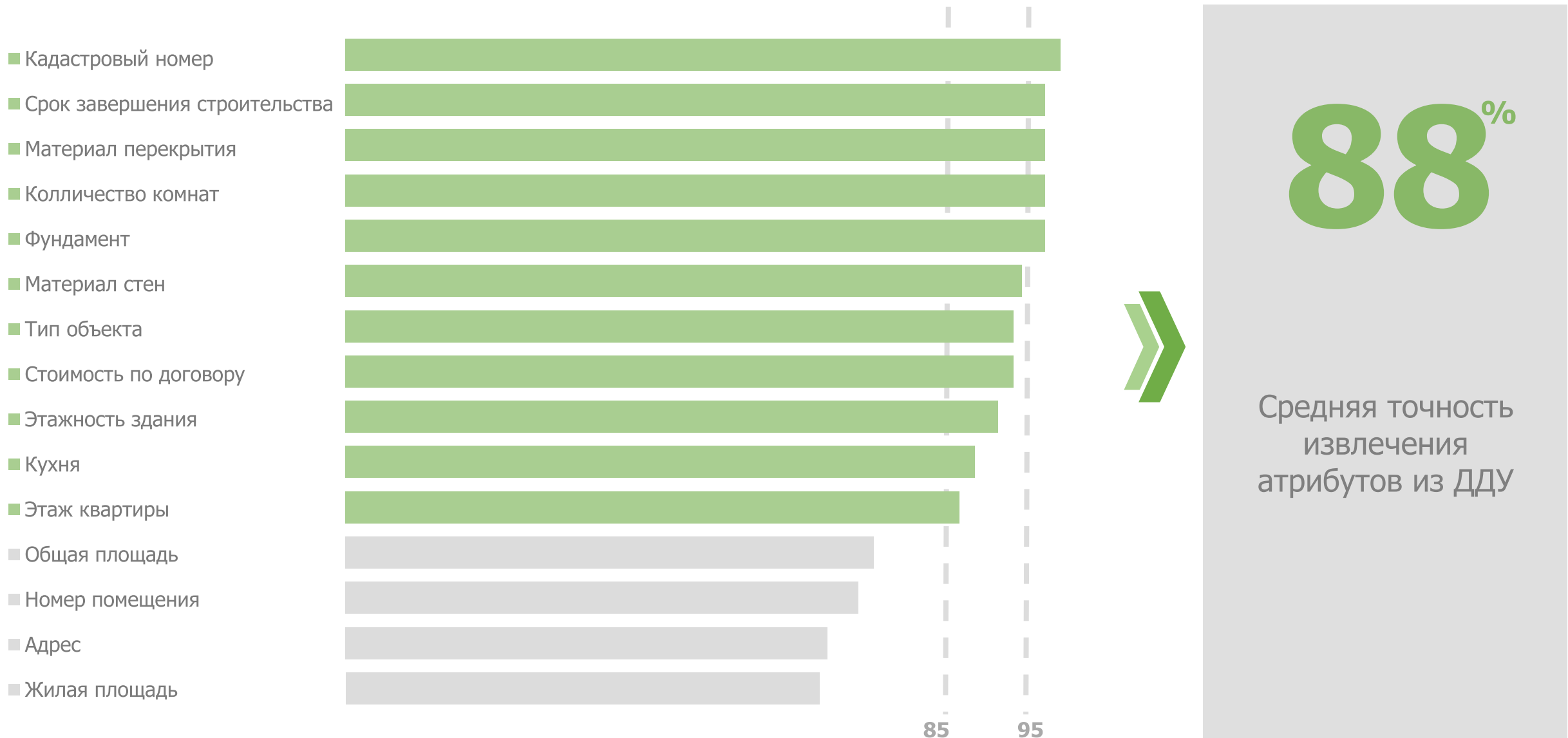


96%

Средняя точность извлечения атрибутов из бухгалтерской документации

* Счет, счет-фактура, акт, акт сверки

Точность извлечения атрибутов из ДДУ*, %



* Машиночитаемый формат, тестовая выборка 92 документа

Эффекты от внедрения DOM.IDP



+31%

Скорость

Оптимизация скорости
обработки документа

Оптимизация ручного труда

Экономия ФОТ при росте клиентского
потока - один менеджер обрабатывает
больше клиентов