open{code:}
открытый код

Библиотеки подпрограмм (SDK) для текстопонимания и текстогенерации на основе технологий машинного обучения

Новизна решения

Онтологический подход - в качестве основополагающей методологии обработки и анализа корпоративной документации, а так же извлечения требуемых данных из них.

Высокая скорость обработки документов за счёт применяемого стека алгоритмов машинного обучения и извлечения информации из корпоративной документации.

Наличие открытых задокументированных интерфейсов интеграции, а также примеров кода интеграции, обеспечит возможность информационного взаимодействия с внешними источниками данных и позволят встраивать библиотеки в корпоративные информационные системы и ресурсы.

Использование собственных методик и правил извлечения данных из корпоративных документов.

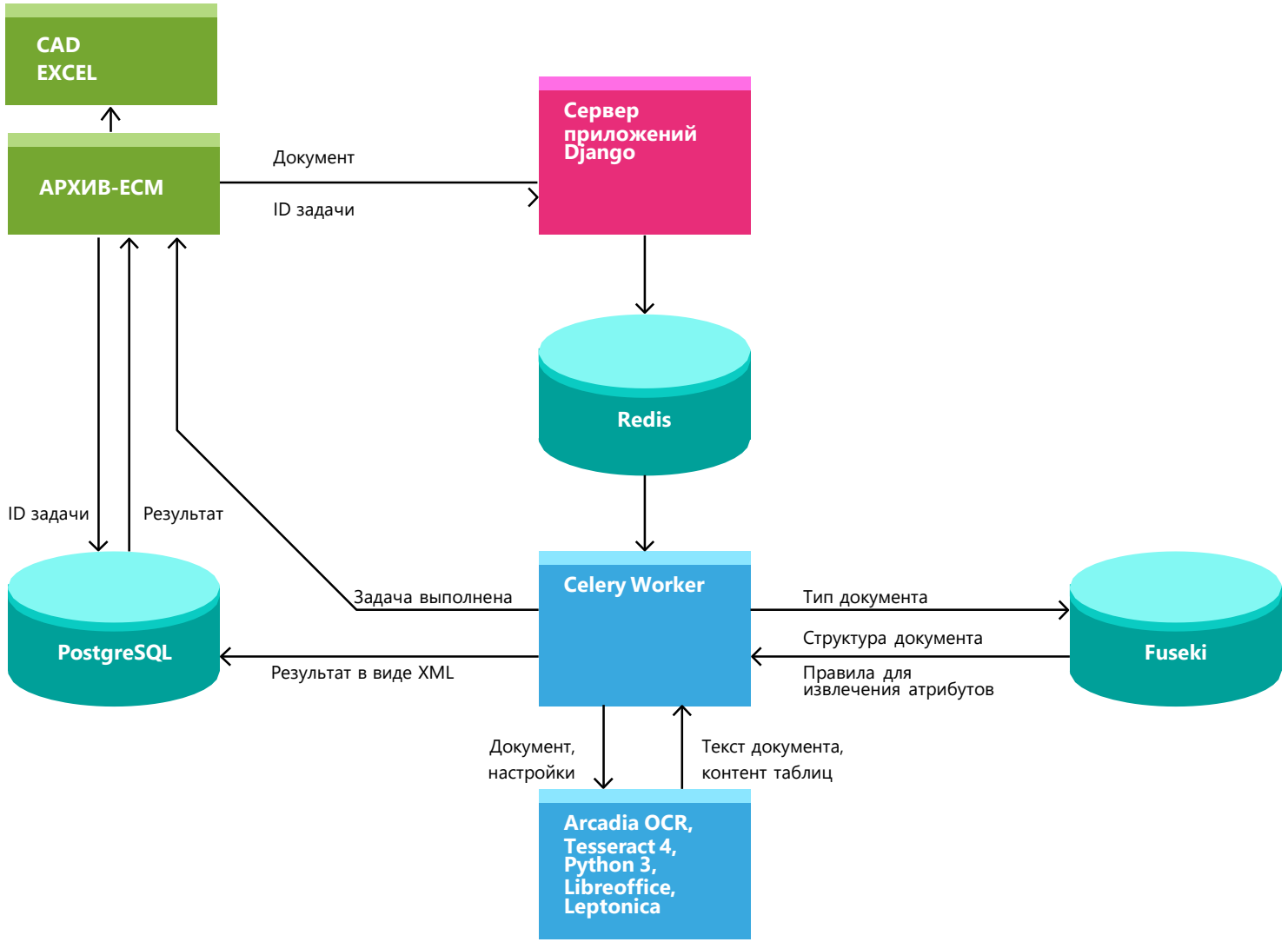
open{code:} открытый код

Основа решения – Собственная разработка компании

SDK входит в состав программного комплекса управления цифровым контентом **«ECM-Интеллект».**

Свидетельство об официальной регистрации программы для ЭВМ №2018663738.

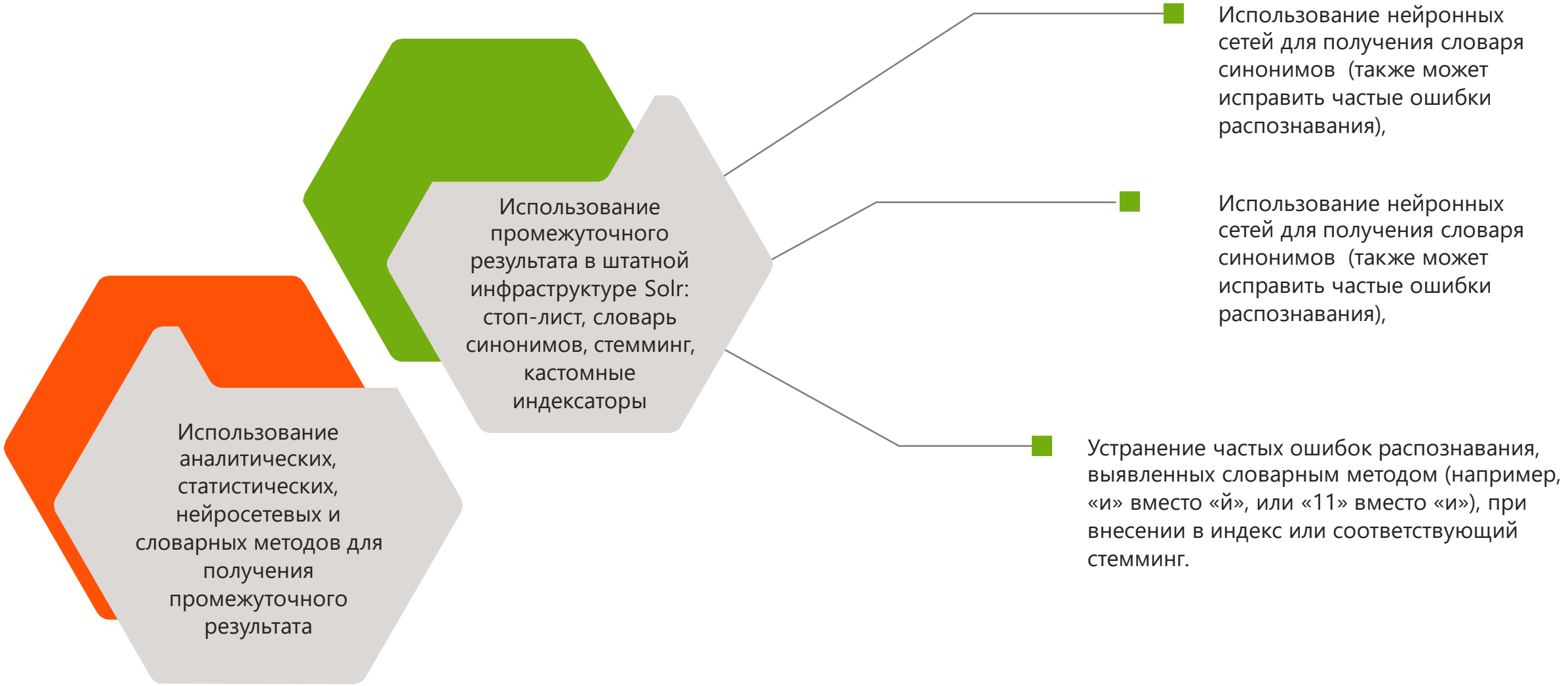
Архитектура решения



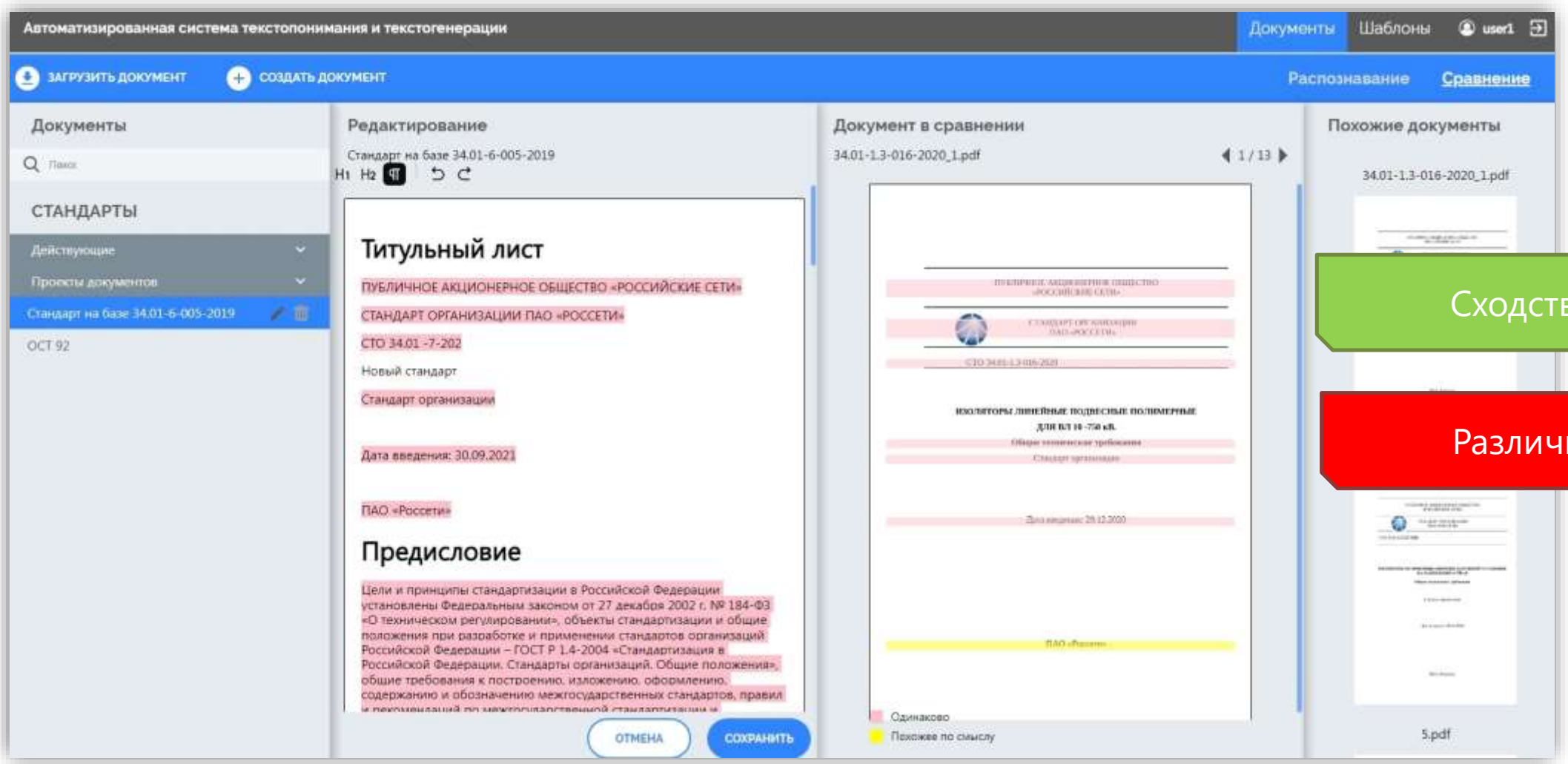
- **CAD** – CAD-система, TeamCentr или любая другая система, которую необходимо обеспечить инструментом текстопонимания и текстогенерации
- **PostgreSQL** – хранилище временных файлов
- **Сервер приложений** – принимает запросы, отправляет документы в базу данных и ставит исполнение в очередь задач
- **Redis** – очередь задач
- **Celery worker** – сервер-исполнитель программного кода, выполняет логику работы для поступающих из очереди (редис) задач
- **Arcadia OCR** – распознавание текста, извлечение данных
- **Fuseki** – СУБД для хранения онтологий.

Уровни реализации

Уровни реализации



Семантическая обработка данных



Семантический поиск

Основа – базы знаний и онтологии.

Здесь релевантность документа запросу определяется семантически, а не синтаксически. В его основе - более точная интерпретация поисковых намерений пользователя.

Это явление многофакторное, обрабатывающее сложные логически обоснованные запросы, которые невозможно решить за счет традиционных видов поиска информации.

«Семантический Анализ документов». Пример