

# Концепция системы подачи заявок на регистрацию объектов интеллектуальной собственности

## Оглавление документа:

1. Контекст проблемы
2. Цели и задачи системы
3. Сценарии использования
4. Функциональные требования
5. Технические(нефункциональные) требования
6. Архитектура системы, схема интеграций
7. База данных для системы
8. Использование AI для поиска, оцифровки документов и построение аналитических моделей
9. Масштабируемость системы и последствия введения её в эксплуатацию

## Контекст проблемы:

1. В патентном ведомстве страны X планируется модернизация системы подачи заявок на регистрацию объектов интеллектуальной собственности и ведения делопроизводства по заявкам.
2. В текущей системе заявку можно подать на бумаге, лично, через представителя или по почте. Объем ежегодно подаваемых в ведомство заявок невелик, из-за сложности заполнения бумажных бланков и высоких пошлин.
3. Делопроизводство по заявке ведется с использованием компьютеров, но без применения автоматизированных информационных систем. Документы (исходящая корреспонденция) готовятся в офисных программных пакетах.

Требуется разработать информационную систему, которая сможет собирать данные об объекте интеллектуальной собственности от заявителя, формировать требуемые документы и отправлять его на согласование разным лицам. Система должна быть масштабируемой, иметь высокий уровень защиты информации, иметь дополнительные возможности, такие как:

1. Поиск в электронном Гос. реестре с использованием AI
2. Оцифровывание документов с бумажных носителей в реестр

## Цели и задачи для системы:

Цель системы: оптимизировать цепочку делопроизводства по регистрации объектов интеллектуальной собственности, используя современные информационные технологии

Задачи системы:

1. Принимать данные о объекте интеллектуальной собственности от заявителя на веб-сервисе ведомства
2. Создавать требуемые документы для регистрации по заданным шаблонам
3. Отправлять данные документы по заданной цепочке сотрудникам компании (Отделу входящей корреспонденции, Отдел экспертизы)
4. Создавать документы для отдела Исходящей корреспонденции
5. Иметь возможность поиска по базе электронного Гос. реестра с использованием AI
6. Иметь возможность загрузки документов, оцифрованных с бумажных носителей, в базу электронного Гос. реестра,

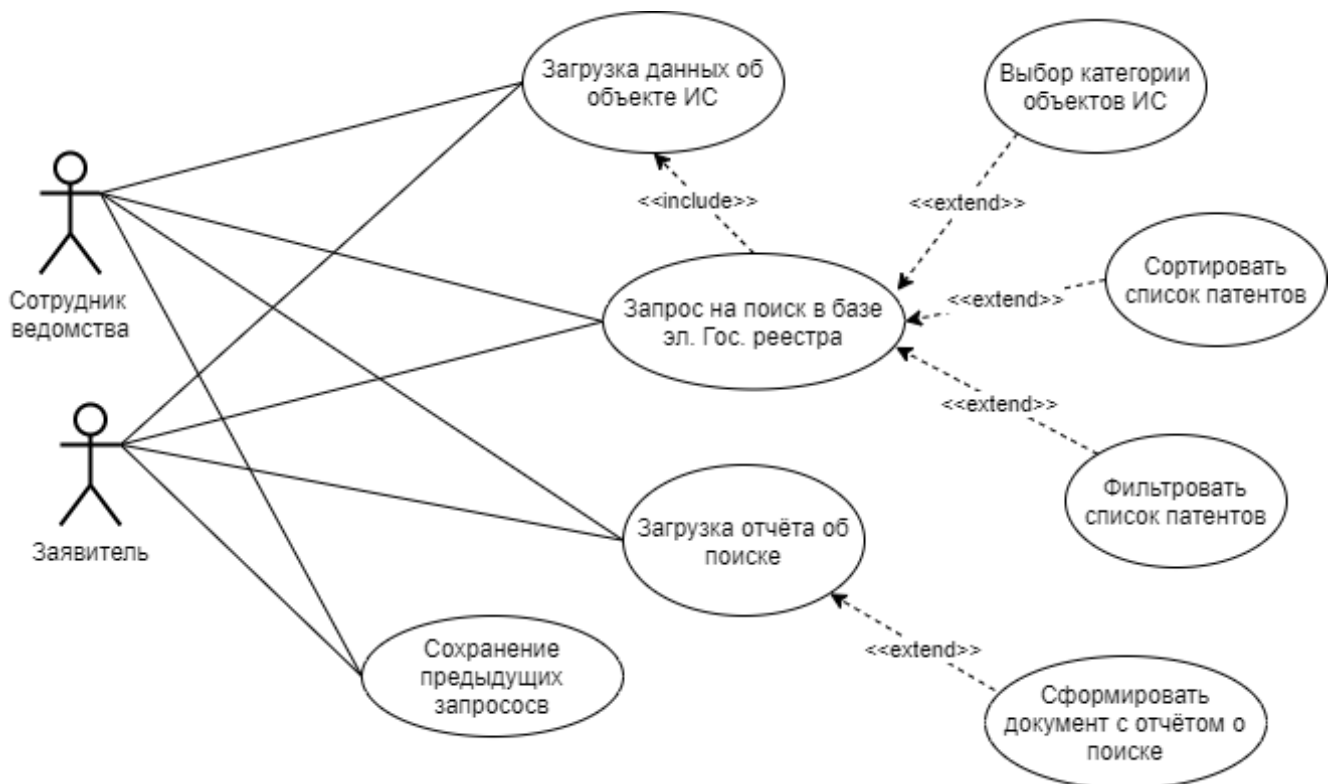
### Сценарии использования:

Основными сценариями использования являются

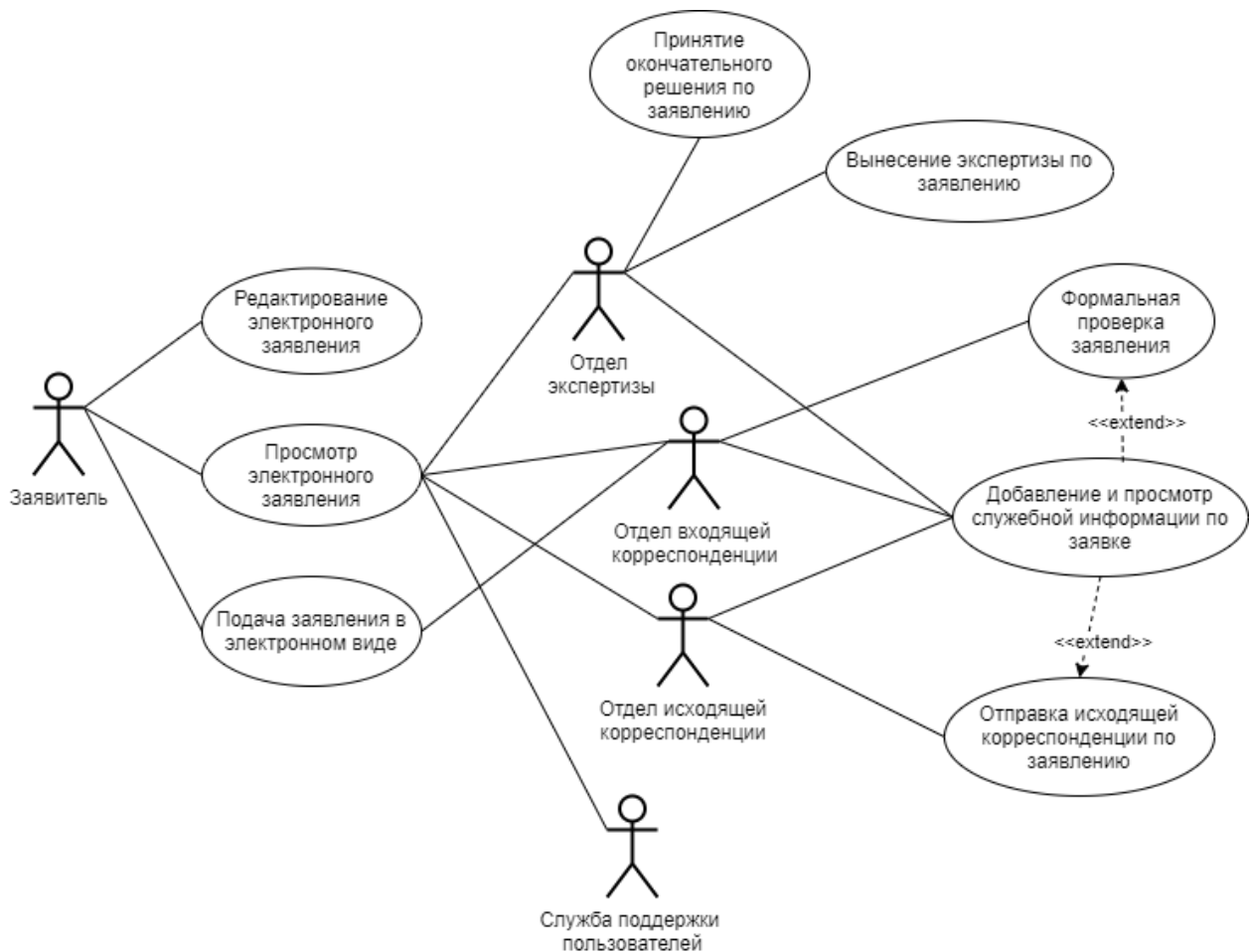
- Поиск в электронном Гос. реестре похожих объектов интеллектуальной собственности
- Создание, редактирование, удаление заявки на регистрацию объекта интеллектуальной собственности
- Отправка и принятие документов от различных отделов на регистрацию объекта интеллектуальной собственности
- Оцифровывание существующих документов в базу данных

Рассмотрим USE CASE для конкретных сценариев:

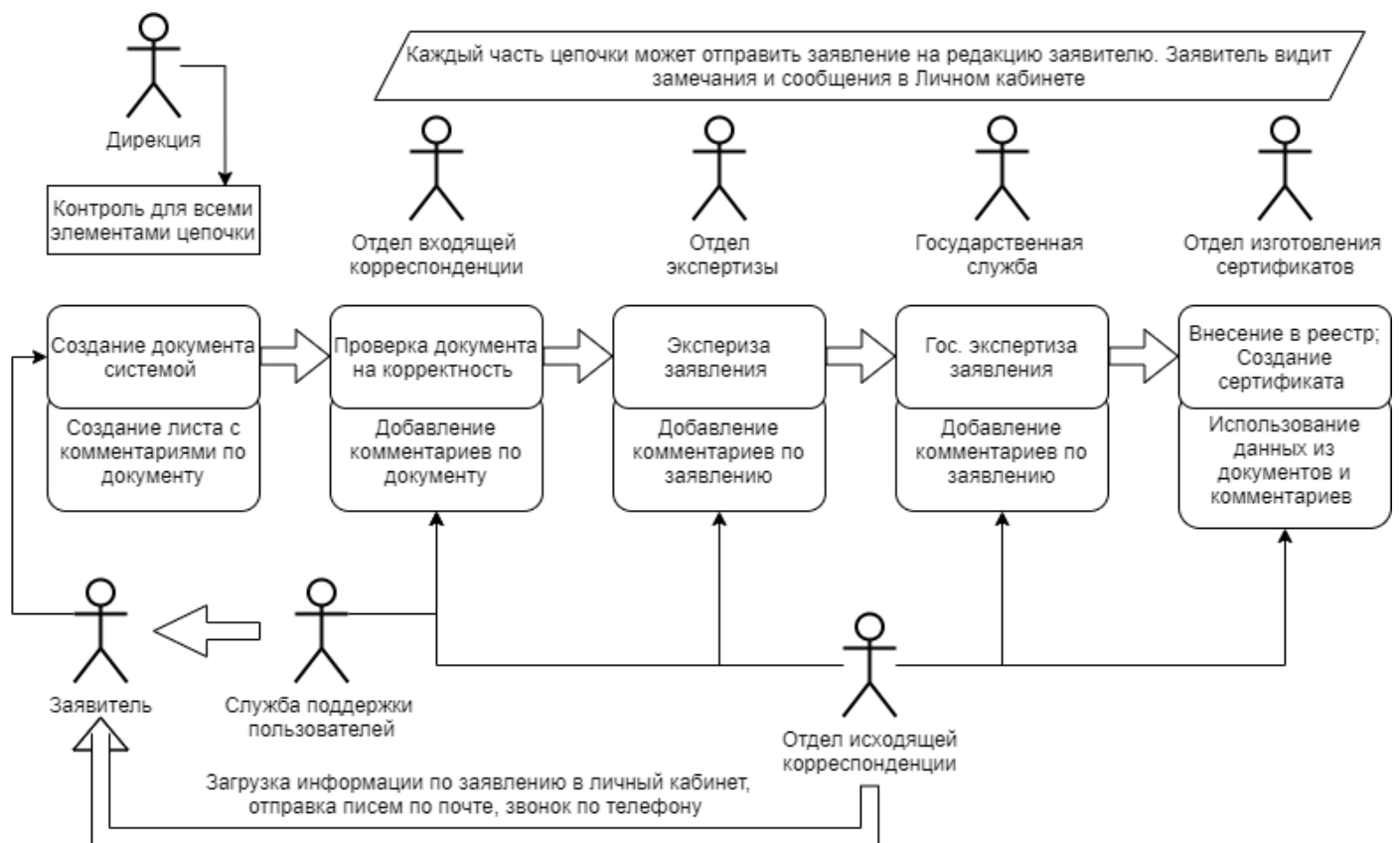
1. Поиск в электронном Гос. реестре



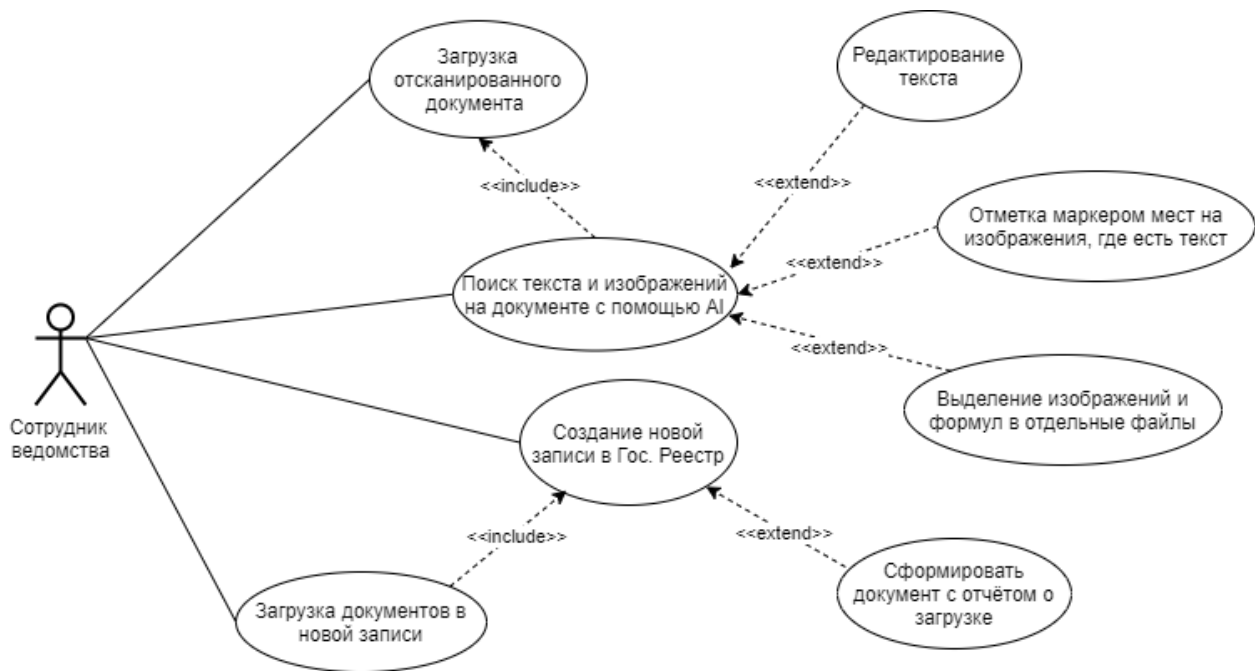
2. Создание, редактирование, удаление заявки



3. Отправка и принятие документов от различных отделов – тут удобнее построить диаграмму процесса



4. Оцифровывание существующих документов в базу данных



## Функциональные требования

Исходя из описанных выше USE CASE можно составить список конкретных функциональных требований:

1. Работа заявителя в системе:
  - Регистрация и авторизация заявителя в системе по номеру телефона и пароля
  - Загрузка информации о себе в систему
  - Создание нового заявления в системе
  - Заполнение информации о заявлении в системе
  - Загрузка изображений и схем по заявлению в систему
  - Редактирование заявления
  - Отправка заявления на формальную проверку
  - Оплата услуг ведомства
2. Работа Отдела входящей корреспонденции в системе:
  - Просмотр входящих заявлений в системе
  - Редактирование документов, сформированных системой
  - Создание новых документов по заявлению
  - Добавление комментариев по документам и заявлению
  - Согласие или отказ по заявлению(формальная проверка)
  - Поиск по электронному Гос. Реестру
3. Работа Отдела экспертизы:
  - Просмотр заявлений в системе
  - Редактирование документов, сформированных системой
  - Создание новых документов по заявлению
  - Добавление комментариев по документам и заявлению
  - Согласие или отказ по заявлению(Экспертиза)
  - Поиск по электронному Гос. Реестру

4. Работа Отдела изготовления государственных сертификатов о регистрации объекта интеллектуальной собственности
  - Просмотр заявлений в системе
  - Добавление комментариев по документам и заявлению
5. Работа Дирекции ведомства:
  - Просмотр заявлений в системе
  - Редактирование документов, сформированных системой
  - Создание новых документов по заявлению
  - Добавление комментариев по документам и заявлению
  - Отслеживание потока документов и заявление
  - Назначение экспертов на проверку документов
  - Отправка уведомлений сотрудникам
6. Работа Финансового отдела:
  - Получение отчётов о заявлениях, транзакциях
  - Оформление счётов на заявителя
  - Добавление комментариев к заявлению
7. Работа Отдела исходящей корреспонденции:
  - Просмотр заявлений в системе
  - Отправление комментариев в личный кабинет заявителя
  - Отправление автоматических почтовых рассылок заявителям
  - Назначение онлайн-встречи
8. Работа любого сотрудника ведомства:
  - Авторизация в системе, как сотрудник
  - Восстановление логина и пароля в сети ведомства

Список функциональных требований к системе может дополняться от потребностей заказчика. Самое важное, что у каждого сотрудника будет свой личный кабинет, в зависимости от его роли в компании, который можно постоянно улучшать, наращивать его функционал для удобства сотрудника

### **Технические(нефункциональные) требования**

Основными требованиями тут будут – скорость работа системы и его безопасность. Требуется быстрая передача документов между отделами ведомства и при этом требуется высокая степень защиты информации. Можно сформулировать следующие требования:

1. Производительность: Система должна быть отказоустойчивой, должна быть работоспособной при полной загрузке ведомства(когда все сотрудники используют систему)
  - Время загрузки документов: не менее 100 MB/sec внутри сети ведомства
  - Время отклика пользователя системы: не более 0,3 секунды
  - Время загрузки интерфейса системы: не более 5 секунд

2. Совместимость: должна поддерживать загрузку данных из базы данных, используемой ведомством и гос. учреждениями, с которыми идёт постоянный обмен данными
3. Удобство использования: интерфейс системы должен быть минималистичным, иметь небольшое кол-во кнопок, иконок, для быстрого обучения пользования ведомством
4. Сопровождаемость: система должна иметь поддержку внесения быстрых изменений, Тех. Обслуживание в год не должно превышать 5% от стоимости его создания

Отдельно стоит рассмотреть защиту данных сотрудников, заявителей и данных о заявлениях. Эта информация может быть использована разными лицами: контрагентами, конкурентами, партнёрами заявителей, что повлечёт серьёзный урон по репутации ведомства. Поэтому предлагаются следующие меры и требования по информационной безопасности:

1. Использование методов криптографической защиты от взлома хакерами: Advanced Encryption Standard
2. Данные о заявлениях должны храниться и передаваться только во внутренней сети. Вне данной сети могут быть переданы только комментарии о заявлении
3. Внутри ведомства должны использоваться аппаратные методы защиты информации: генераторы шума, сетевые фильтры, сканирующие радиоприемники
4. Нужно использовать также программный подход: использовать современные антивирусные пакеты
5. Организационный метод: данные должны использоваться тем кругом лиц, кому это действительно необходимо. В нашей системе – данные о заявлении могут просматривать только сотрудники, имеющие непосредственное отношение к проверке заявления.

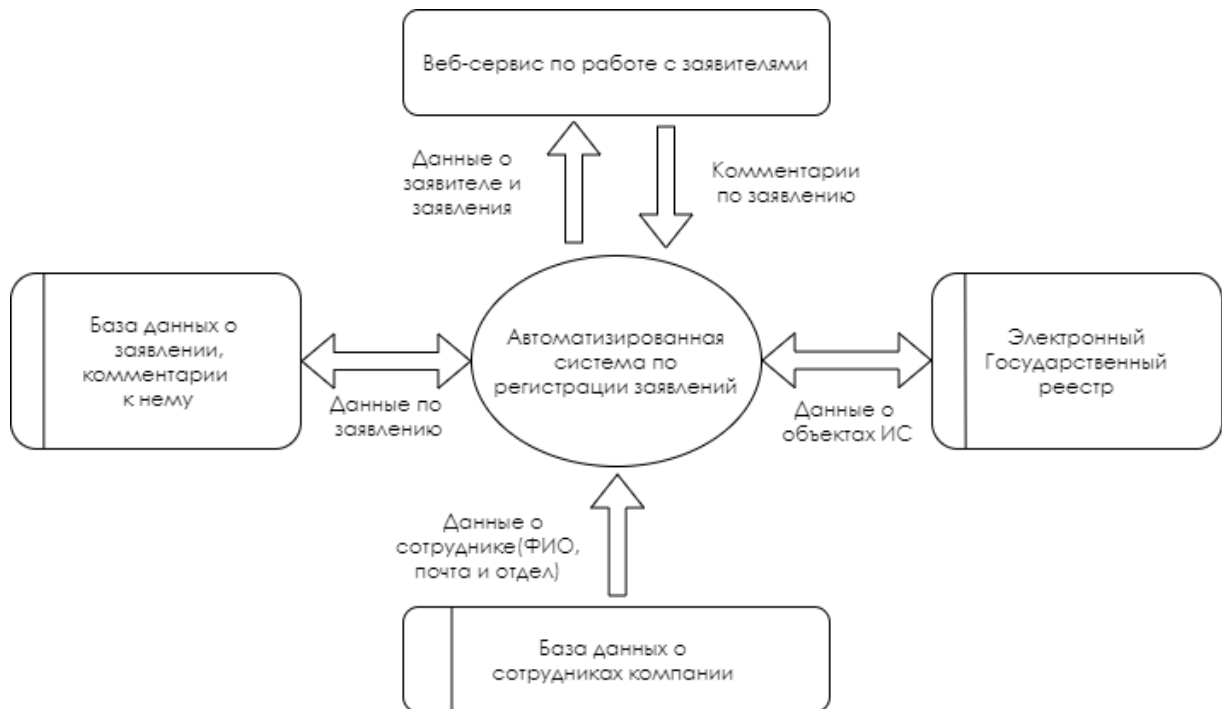
Таким образом мы можем устранить или уменьшить риск следующих угроз:



Также требуется использовать электронную подпись для дополнительной защиты прав заявителей и всех участников процесса. Можно использовать сертификат X.509 или используя готовые решения: «Криптопро CSP», Signal-com CSP, «Лисси CSP», Vipnet CSP.

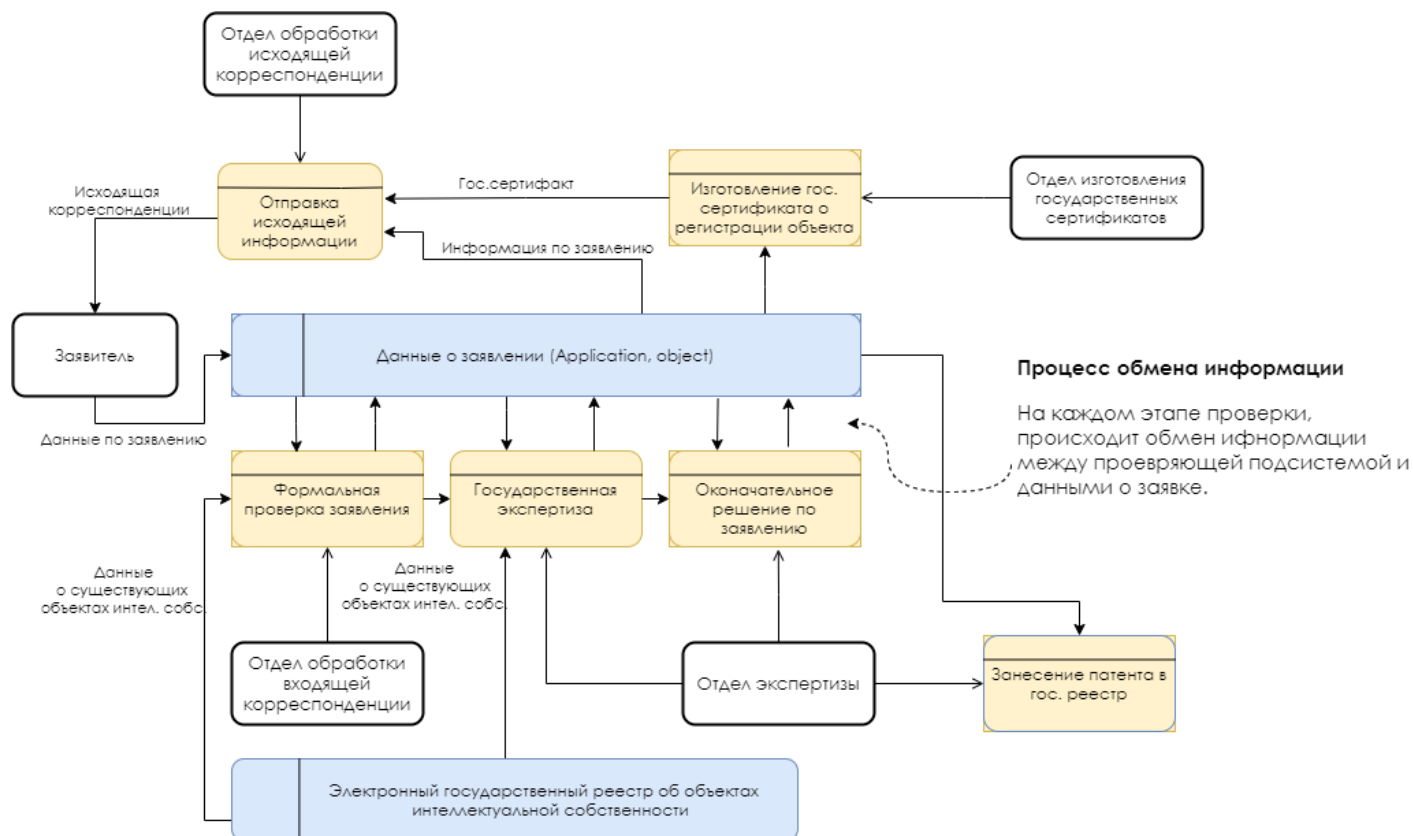
### Архитектура системы, схема интеграций

Система должна иметь высокие нефункциональные требования, поэтому её архитектура должна позволять быстро и безопасно передавать информацию между её частями. Составим концептуальную схему передачи информации и интеграция с другими сервисами :

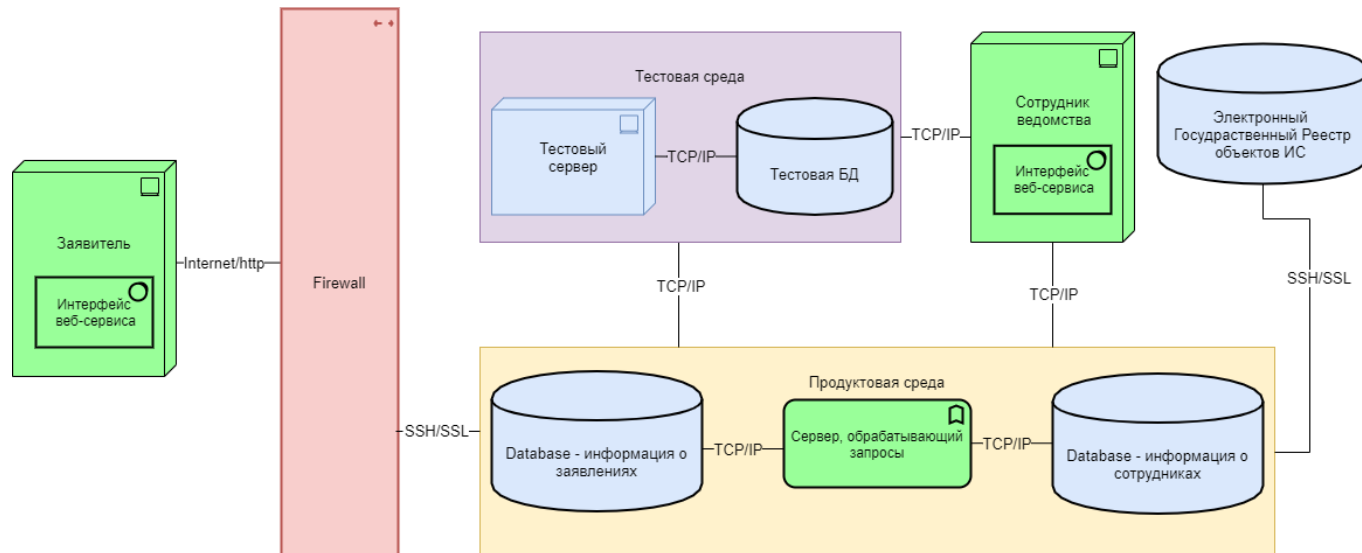


Таким образом мы должны обеспечить быстрый и безопасный способ передачи информации по всем 4 каналам связи. 3 из них находятся внутри сути ведомства и здесь не требуются высоких требований к защите информации, ведь злоумышленники не смогут к ней подключиться, учитывая только физический доступ к данным. Поэтому тут отдаётся предпочтение скорости, которую можно повысить используя высокотехнологичные оптоволоконные кабели. Остаётся канал связи: Система – Электронный Гос. Реестр. По данному каналу с одной стороны идёт общедоступная информация о существующих патентах, а с другой стороны идут заявки на регистрацию новых объектов интеллектуальной собственности.

Отсюда можно рассмотреть также диаграмму потока данных.



Также можно рассмотреть диаграмму развёртывания:



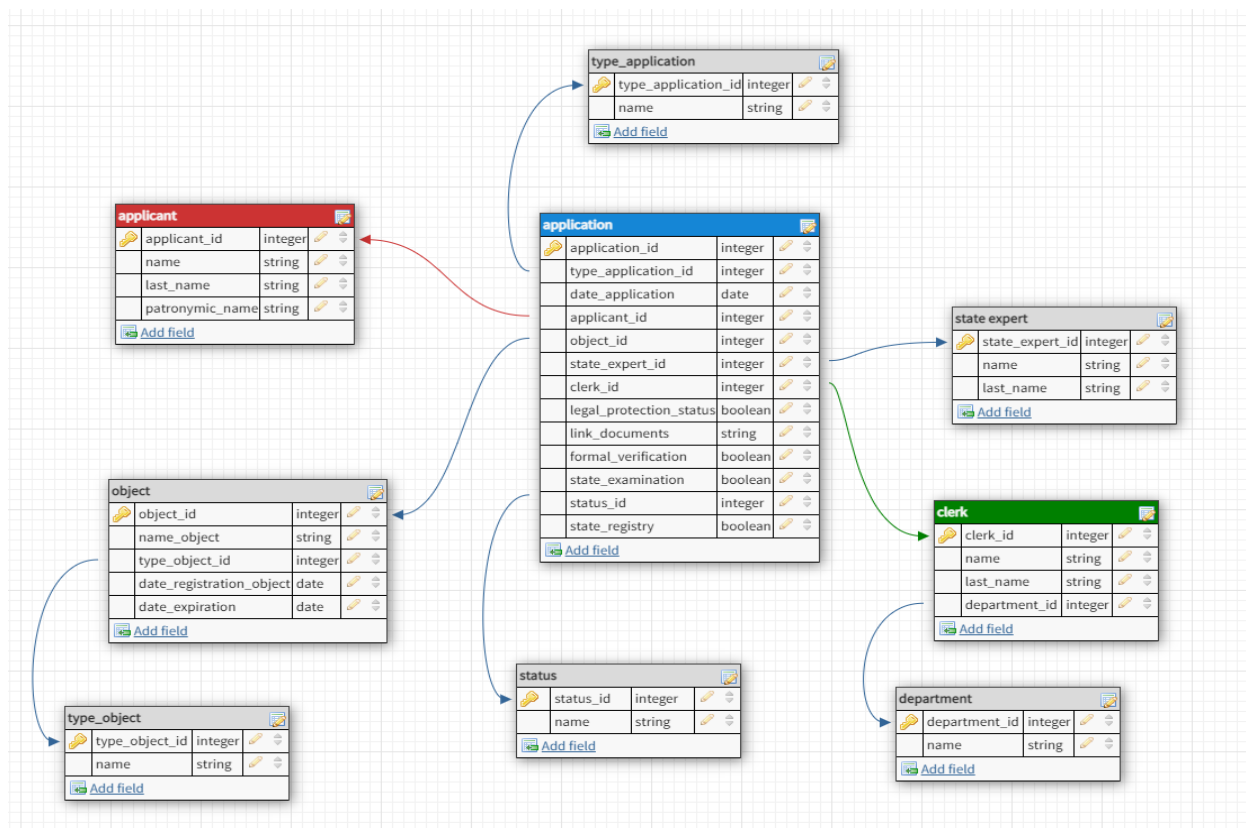
В данной диаграмме используются определённые каналы связи, обеспечивающие высокие параметры скорости передачи данных, имеющих высокую защиту от взлома и кражи данных. Также была добавлена тестовая среда, в которой можно будет пробовать новый функционал системы перед тем как использовать его в продуктовой версии. Это снизит вероятности сбоев и простоя сервиса.



В данный момент, предполагается использование отдельной базы данных для хранения данных о заявителях и о их заявлениях в систему. Но в будущем можно создать единую БД, в которой будет храниться вся нужная информация, что снизит нагрузку на всю систему в целом.

### База данных для системы

Для работы с данным сервисом была создана БД, в которой можно хранить информацию о заявлениях и заявителях



В также были создана таблицы для хранения информации о сотрудниках, работающих с данным заявлением, для дальнейшей интеграции двух БД

### Использование AI для поиска и построение аналитических моделей

В нашем случае у нас есть множество способов использования методов искусственного интеллекта для оптимизации бизнес-процессов. В первую очередь множество ускорить процесс поиска похожих записей в Гос. реестре по заданному тексту или изображению. Для данной задачи уже существуют множество разных готовых методов. Так, для поиска по тексту можно использовать методы решения задачи Paraphrase Identification в NLP:

- Abstract Meaning Representation (AMR)
- Microsoft Research Paraphrase Corpus
- Bilateral multi-perspective matching
- Stack augmented Parser-Interpreter Neural Network

Методика Microsoft доступна бесплатно и в целом на рынке существуют готовые решения от компаний, как ABBYY, которые занимаются данной задачей на рынке больше 10 лет, используя новейшие методы машинного обучения

Также очень важно проводить поиск похожих изображений. Данная задача называется - Content-Based Image Retrieval. Существует также множество готовых методов решения данной задачи:

- VGG16
- ResNet50
- Inception
- GoogLeNet
- FIRE
- MIFile
- Anaktisi
- И многие другие

Существующих методов достаточное количество, среди них есть открытые, которые можно использовать без покупки лицензии, выложен программный код в открытый доступ. Есть множество готовых решений от Яндекс и Google, которые можно интегрировать в систему.

Также стоит задача по оцифровке существующих документов с бумажных носителей в электронную базу. Можно отсканировать существующие документы, а затем сканировать текст в них используя методы OCR(Optical character recognition). Такие решения есть в открытом доступе от многих компаний – Adobe, ABBYY и так далее.

Также можно использовать методы машинного обучения для прогнозирования загруженности системы и всего ведомства в целом. У нас есть исторические данные потоку заявок и мы можем построить прогноз, каким будет поток через определенное время. Есть огромное множество алгоритмов, методов решения данной задачи:

1. Регрессионные модели прогнозирования
2. Авторегрессионные модели прогнозирования (ARIMAX, GARCH, ARDLM)
3. Модели экспоненциального сглаживания (ES)
4. Модель по выборке максимального подобия (MMSP)
5. Модель на нейронных сетях (ANN)
6. Модель на цепях Маркова (Markov chains)
7. Модель на классификационно-регрессионных деревьях (CART)
8. Модель на основе генетического алгоритма (GA)
9. Модель на опорных векторах (SVM)

10. Модель на основе передаточных функций (TF)

11. Модель на нечеткой логике (FL)

12. И много других методов

Но у каждого из способов есть свои недостатки и преимущества, своя область применения, свои ограничения по данным, по типу данных. Поэтому не существует универсального решения, но мы можем подобрать наиболее релевантные алгоритмы для нашей задачи. Наша модель должна лучше всего удовлетворять следующим метрикам:

1) MAPE – средняя абсолютная ошибка в процентах

2) MAE – средняя абсолютная ошибка

Это самые основные и популярные метрики в прогнозировании временных рядов. Также надо учитывать, что мы предсказываем дискретную величину, что сразу отбрасывает множество методов. Так, лучше всего подходит метод - Авторегрессионные модели прогнозирования (ARIMAX). Он хорошо изучен, есть множество модификаций, можно качественно настроить под свои задачи и при этом – сама реализация несложная, всю задачу может выполнить один человек. Для каждого вида нужно будет выбрать свою модель, свои настройки, возможно особенную модификацию.

### **Масштабируемость системы и последствия введения её в эксплуатацию**

Благодаря тому, что система не ограничена числом заявлений, поступающих в момент времени, можно постоянно увеличивать мощности сервера, от требуемых потребностей сотрудников. Благодаря тому, что вся система будет расположена в офисе ведомства и данные будут передаваться по внутренним каналам – будет высокая скорость передачи информации между её узлами. Таким образом это всё позволит значительно увеличить поток клиентов для ведомства, этому также поспособствует скидка 30% на пошлины регистрации объектов интеллектуальной собственности за счёт электронной подачи всех документов в Государственную службу. А вся подача заявлений в дистанционном формате позволит принимать их со всей страны, во всех регионах, а также из дружественных государств, ведь теперь не потребуется визит офиса ведомства для согласования или подписывания договоров, так как есть система подписывания с помощью электронной подписи.

Это система будет чем-то похожа на процесс документооборота, в которой документы будут автоматически отсылаться определённым лицам.

Это позволяет значительно снизить потребление бумаги, а также снизить издержки, связанные с бумажными носителями, такими как:

- Недолговечность
- Могут потеряться
- Сложно транспортировать в больших масштабах
- Могут быть уничтожены, в последствии пожара или других аварий в офисе.

И все проблемы будут решены с помощью автоматического решения по регистрации объектов интеллектуальной собственности