

# КОРПОРАТИВНЫЙ ПОИСК

Каталог ИИ-решений

# ЧТО ВЫ ПОЛУЧАЕТЕ С ИСПОЛЬЗОВАНИЕМ КОРПОРАТИВНОГО ПОИСКА



Умный поиск без потерь  
рабочего времени



Полную и актуальную информацию  
для принятия решений



Единое информационное  
пространство



Удобный доступ к накопленным  
компанией знаниям

## ВЕБ-ИНТЕРФЕЙС ПОИСКОВОЙ СИСТЕМЫ

- ✓ Пользователь вводит запрос и получает выдачу в интерфейсе
- ✓ Есть возможность из поиска перейти в источник

## ВИДЖЕТ ПОИСКА В КОРПОРАТИВНЫЕ ВЕБ-СИСТЕМЫ (ПОРТАЛЫ, ПОЧТОВЫЙ КЛИЕНТ И ДР.)

- ✓ Пользователь вводит запрос и получает выдачу в интерфейсе корпоративной системы
- ✓ Возможность перехода пользователя в веб-интерфейс NES для использования расширенных инструментов

## API-ИНТЕРФЕЙС

- ✓ Передача поисковой выдачи с учетом фильтров, прав доступа, атрибутов, ссылки на источник), семантической аналитики (похожие, ключевые слова, выделенные объекты, изменения по документам и др.)

## 01 ПОИСК ПО СОДЕРЖАНИЮ И АТТРИБУТАМ

Поиск в контенте, метаданных и значениях атрибутов документа (PDF, DJVU, PPTX и др.)

## 02 ФИЛЬТРАЦИЯ РЕЗУЛЬТАТОВ ПОИСКА

Динамическое построение фильтров с учетом результатов поиска и прав пользователя

## 03 ИСПРАВЛЕНИЕ ЗАПРОСА

Автоматическая корректировка орфографии и опечаток в поисковом запросе пользователя

## 04 НАСТРОЙКА РЕЛЕВАНТНОСТИ

Администратор системы может регулировать веса документов в выдаче

## 05 АВТОПОДСКАЗКИ

Когда пользователь начинает вводить запрос, ему предлагаются возможные варианты его завершения – слово, словосочетание или заголовок документа

## 06 ЯЗЫК ЗАПРОСОВ

Использование логических операторов для более точного формулирования поискового запроса

# ВОЗМОЖНОСТИ СИСТЕМЫ, ЕСЛИ ПОЛЬЗОВАТЕЛЬ НЕ ЗНАЕТ, ЧТО ИЩЕТ

## 01 РАСШИРЕНИЕ ПОИСКОВОГО ЗАПРОСА

В механизм входит:

- ✓ Поиск по синонимам
- ✓ Поиск по аббревиатурам и терминам

База синонимов может содержать общую лексику, лексику узкой предметной области, а также специфичные термины и сокращения, используемые в компании

## 02 ПОИСК ПО СМЫСЛУ

Поиск документов со схожим смыслом с использованием технологии векторного поиска

## 03 ПОИСК ПО ТЕМАТИЧЕСКИМ ГРУППАМ

- ✓ Выделение тематических групп на корпус документов заказчика
- ✓ Редактирование состава тематических групп и наименований
- ✓ Перестраивание групп при добавлении новых данных

## 04 КАТАЛОГ ДАННЫХ

- ✓ Создание иерархической структуры базы знаний
- ✓ Ассоциация документов со структурой категорий по расписанию

## 05 ПОИСК ПО ТЕГАМ

Присвоение документам текстовых меток по набору правил





## 06 ПОИСК ПО КЛЮЧЕВЫМ СЛОВАМ

Выделение из текста слов, которые дают высокоуровневое описание содержания документа

## 07 ПЕРСОНАЛИЗАЦИЯ ПОИСКА

Поисковая выдача зависит от знаний об отдельном пользователе (личные данные, приоритеты, интересы и т. д.). С каждым запросом данные о предпочтениях пользователя накапливаются и поиск становится точнее

## ИСТОЧНИКИ

-  Документы пользователя
-  СЭД, ERP-системы
-  Корпоративные хранилища данных
-  Внешние источники

## РОЛЕВАЯ МОДЕЛЬ

Служба каталогов  
**LDAP/Active Directory**

## ИНТЕЛЛЕКТУАЛЬНАЯ ПЛАТФОРМА КОРПОРАТИВНОГО ПОИСКА

### Семантический поиск

- ✓ Векторный поиск
- ✓ Поиск неявных связей (векторное сравнение, текстовые пересечения, обогащение связанными внешними данными)




### Текстовая аналитика

Извлечение именованных сущностей и ссылок, поиск семантических связей, саммаризация, поиск противоречий и проверка на соответствие шаблону

### Структурирование данных

- ✓ Построение и наполнение каталога на основе моделей кластеризации;
- ✓ Выделение из текста ключевых слов, которые дают высокоуровневое описание содержания документа


## ИСТОЧНИКИ

-  API-интерфейс к поисковой платформе
-  Веб-виджет поисковой строки для корпоративных систем и почты
-  Основной веб-интерфейс

# 1. ПОЛУЧЕНИЕ ДАННЫХ ИЗ ИСТОЧНИКОВ


## ПОИСК ПО СМЫСЛУ

- ✓ Готовые коннекторы для подключения корпоративных систем
- ✓ Самостоятельное подключение источников с помощью стандартизированного API-интерфейса

 Актуально для enterprise систем типа SharePoint, IBM Lotus, IBM FileNet


## РОБОТЫ-КРАУЛЕРЫ

- ✓ Краулеры для сбора данных с порталов и систем, а также Apache Nutch или самописных «пауков» по спецификации источников

 Актуально для внешних систем и типовой индексации страниц порталов без коннекторов


## ВСПОМОГАТЕЛЬНЫЕ СЕРВИСЫ

- ✓ КШД или DFS, временное хранилище для передачи информации

 Интеграция через файлы с выгрузками в заданном формате или копии витрин данных

## Веб-сервисы

- ✓ Совместная разработка веб-сервисов для передачи и приема данных от системы источников в Корпоративный поиск

 Актуально для кастомной интеграции. Подойдет при наличии сложных полей

## 2. СЕМАНТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ И ПОСТРОЕНИЕ ПОИСКОВОГО ИНДЕКСА

### Построение поискового индекса

- ✓ Построение эмбединга - семантического вектора документа
- ✓ Токенизация - выделение основных структурных единиц речи
- ✓ Вычисление наиболее часто встречающихся N-грамм
- ✓ Фильтрация лемм - исключение незначимых словарных конструкций

### Текстовая аналитика

- ✓ Выделение именованных сущностей для дополнительной атрибуции
- ✓ Выделение ссылок для выявления связей между документами и обогащения внешними данными (НПА)
- ✓ Семантическое сравнение и поиск текстовых пересечений
- ✓ Аннотирование документов
- ✓ Аннотирование поиск противоречий и проверка на соответствие шаблону

## 3. ПРЕДОБРАБОТКА ДАННЫХ

### ИЗВЛЕЧЕНИЕ КОНТЕНТА

Разархивация (ZIP, RAR, 7Z и др.) и извлечение текстового слоя из:

- ✓ Документов Microsoft (doc, docx, xls,xlsx, ppt, pptx и др.)
- ✓ Текстовых документов открытого формата (odt)

### РАСПОЗНАВАНИЕ И ИЗВЛЕЧЕНИЕ ТЕКСТОВОГО СЛОЯ

Распознавание документов без текстового слоя (pdf, djvu, jpeg)

- ✓ Сканы рукописных файлов, документы, чертежи, схемы

### Извлечение метаданных из документа

Извлечение мета-данных файлов (дата создания, путь к файлу и др.)

## 4. СТРУКТУРИРОВАНИЕ ДАННЫХ

### Построение поискового индекса

- ✓ Маппинг мета-данных файлов на атрибуты поиска
- ✓ Маппинг атрибутов интегрируемых систем на атрибуты поиска
- ✓ Построение и наполнение каталога на основе моделей кластеризации или набора правил
- ✓ Выделение из текста ключевых слов, которые дают высокоуровневое описание содержания документа
- ✓ Тегирование

## 5. ФОРМИРОВАНИЕ ПОИСКОВОЙ ВЫДАЧИ



Пользователь вводит запрос

СИСТЕМА

### Обрабатывает запрос, который содержит:

- ✓ Запрос пользователя
- ✓ Значения фильтров и тегов
- ✓ Синонимы и корпоративную лексику
- ✓ Логические операторы

### Формирует поисковую выдачу:

- ✓ Применение векторного поиска (поиск по смыслу)
- ✓ Маппинг на права пользователя

### Ранжирует результаты поиска в зависимости от:

- ✓ Весов атрибутов, источников
- ✓ Релевантности для пользователя (история, организация, др. параметры)

## 01 КОНТРОЛЬ ДОСТУПА

NES интегрируется с системами управления доступом, позволяя предоставлять его только авторизованным пользователям на основе их роли

## 02 ЛОГИРОВАНИЕ И МОНИТОРИНГ

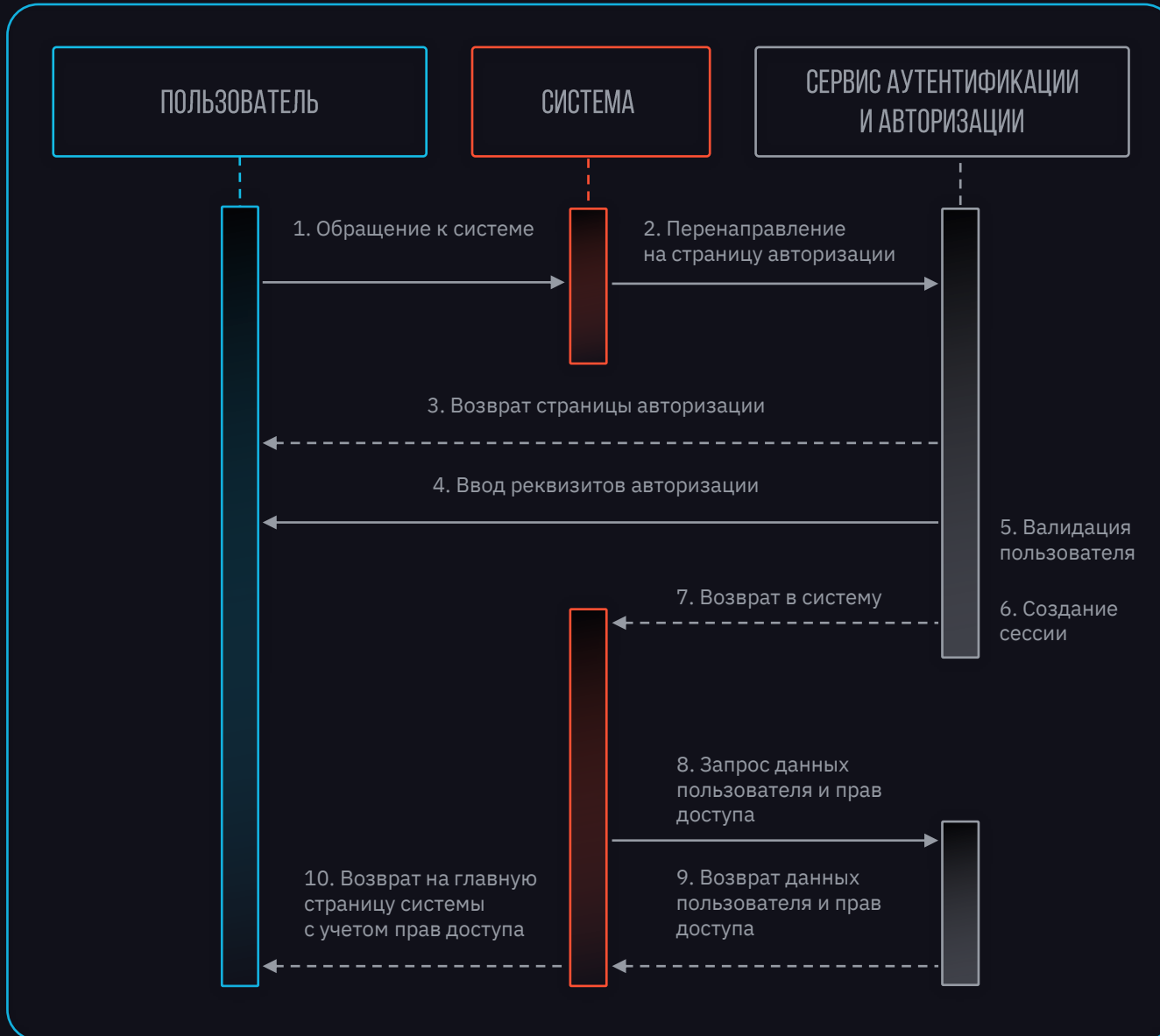
Решение ведет журнал всех поисковых запросов и действий пользователей, что позволяет отслеживать подозрительную активность и выявлять потенциальные угрозы безопасности

## 03 ШИФРОВАНИЕ ДАННЫХ

Интеллектуальный поиск использует методы шифрования для защиты данных как при хранении, так и при передаче, обеспечивая их безопасность от несанкционированного доступа

## 04 ПОЛИТИКИ БЕЗОПАСНОСТИ

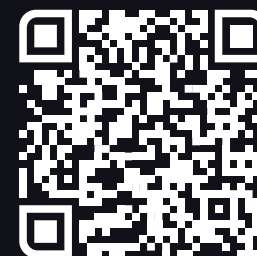
NES настраивается с учетом корпоративной политики безопасности, обеспечивая соответствие стандартам и нормативным требованиям



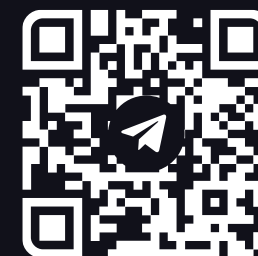
## Варианты реализации сервисов аутентификации и авторизации:

- ✓ Внутренний сервис аутентификации Корпоративного поиска на основе Kerberos-аутентификации
- ✓ Keycloak — система, которая реализует технологию единого входа с возможностью управления доступом
- ✓ LDAP-совместимые службы каталогов, у которых есть возможность интеграции для авторизации и использования групповых политик (например Active Directory Domain Services)
- ✓ Сторонние сервисы, которые имеют возможность аутентифицировать и авторизовывать по REST API

# СПАСИБО ЗА ВНИМАНИЕ!



ФЦПРИИ.РФ



t.me/fcprii