

ООО «Преферентум»

Поисково-аналитическая система
«АйТи-Поиск»

Описание функциональности и архитектуры

Листов 10

Москва

2022

Оглавление

1. ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ	3
2. ВВОДНАЯ ИНФОРМАЦИЯ	5
2.1. НАЗНАЧЕНИЕ ДОКУМЕНТА.....	5
2.2. ОПИСАНИЕ РАССМАТРИВАЕМОГО ПО	5
3. НАЗНАЧЕНИЕ И ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ ПО «АЙТИ-ПОИСК».....	6
4. ОПИСАНИЕ РАБОТЫ СИСТЕМЫ.....	7
5. ВЗАИМОДЕЙСТВИЕ АЙТИ-ПОИСК С ВНЕШНИМИ ИС	7
6. ОПИСАНИЕ КОНЦЕПТУАЛЬНОЙ АРХИТЕКТУРЫ	7

1. Термины и определения

В разделе (см. таблица ниже) представлены термины и определения, а также сущности, с которыми работает Поисково-аналитическая система «АйТи-Поиск».

Термин	Описание
Алгоритмическое правило	Жесткие алгоритмы, настройка выполняется разработчиком или администратором
Аннотация	<p>Информационное сообщение определенного формата и структуры, сформированное в виде примечания к тексту документа.</p> <p>Аннотация:</p> <ul style="list-style-type: none"> Указывает на место в документе, где сработало (должно было сработать) правило проверки текста либо правило извлечения именованной сущности (факта). Несет в себе информацию о том, как сработало (должно было сработать) правило или какой факт был извлечен (должен был быть извлечен). <p>В документе будет применяться терминология в рамках нижеперечисленных видов классификации аннотаций.</p> <p>По природе создания аннотации можно разделить:</p> <ul style="list-style-type: none"> На автоматические, созданные ПО при отправке документа на проверку. На созданные вручную экспертом при проверке результатов автоматической обработки документа. <p>По отношению к типу документа аннотации можно разделить:</p> <ul style="list-style-type: none"> На аннотации извлечения именованных сущностей (фактов). На аннотации, созданные по результатам срабатывания интеллектуальных правил проверки. <p>По отношению к проверяемому тексту аннотации можно разделить:</p> <ul style="list-style-type: none"> На информационные – аннотации, которые не производят корректировку текста проверяемого документа. На корректирующие – аннотации, которые могут автоматически изменить текст проверяемого документа.
Аннотирование	Подтверждение или опровержение Экспертом результатов срабатывания интеллектуальных правил
АРМ	Автоматизированное рабочее место
АС	Автоматизированная система
БД	База данных
ГОСТ	Государственный стандарт
ИС	Информационная система
Интеллектуальные правила	Жесткие алгоритмы, настройки индуцируются системой на основе примеров, задаваемых экспертом
ЛВС	Локальная вычислительная сеть
ЛНА	Локальные нормативные акты
Набор правил	Несколько Правил, объединенных в процесс, исполняемый в модуле «АйТи-Поиск». Набором правил можно управлять (конфигурировать Набор). Например, включать или выключать работу отдельных Правил Набора или устанавливать значения параметров.
НПД	Нормативно-правовые документы
Обратная связь	Информация, внесенная экспертами по результатам проверки задания: решения по срабатыванию правил и предложения новых правил
ОЗУ	Оперативное запоминающее устройство
ООО «Преферентум»	Вендор решения «Поисково-аналитическая система «АйТи-Поиск»»
ОС	Операционная система

Термин	Описание
ПО	Программное обеспечение
ПО «АйТи-Поиск»	Поисково-аналитическая система «АйТи-Поиск»
Правило	Реализованный в виде программного кода интеллектуальный алгоритм, осуществляющий проверку текста на соответствие заданным критериям. Примерами критерия могут быть: наличие или отсутствие определенной информации (Фактов) в документе, соответствие или несоответствие проверяемого документа определенным требованиям: типовой форме или структуре, корректность употребления терминов и выражений, отсутствие ссылок на утратившие силу ГОСТы. Если в результате применения правила выявлены искомые факты или обнаружены ошибки в документе, правило формирует аннотации.
Рейтингование правил	Изменение рейтинга правила для целей проверки допустимости применения правила
СУБД	Система управления базами данных
ТЗ	Документ типа «Техническое задание»
ТКП	Документ типа «Технико-коммерческое предложение»
Факт	Смысловое понятие, выделяемое в тексте документа. Примерами фактов могут быть: наименование, телефон, e-мэйл организации, ФИО руководителя и т.д. Синонимы – атрибут, реквизит, сущность.
AD	Active Directory, службы каталогов корпорации Microsoft для операционных систем семейства Windows Server
API	Программный интерфейс приложения (от англ. application programming interface)
GUI, UI	Графический пользовательский интерфейс (англ. graphical user interface, GUI)
IIS	Internet Information Services, проприетарный набор серверов для нескольких служб Интернета от компании Microsoft
IP-адрес	Уникальный сетевой адрес узла в компьютерной сети, построенной на основе стека протоколов TCP/IP
RabbitMQ	Платформа, реализующая систему обмена сообщениями между компонентами программной системы на основе стандарта AMQP.
REST	Архитектурный стиль взаимодействия компонентов распределённого приложения в сети, сокращенно от Representational State Transfer.
SMB	Сетевой протокол прикладного уровня для удалённого доступа к файлам и другим сетевым ресурсам (от англ. Server Message Block)
SofA	Source of Analysis, источник анализа. Объект данных сложной структуры, содержащий проверяемые набором правил документы в универсальном структурном представлении и набор аннотаций и реквизитов, выделяемых в процессе обработки документов набором правил.
URL	Единообразный определитель местонахождения ресурса (от англ. Uniform Resource Locator)
Web	Программное обеспечение или оборудование, обеспечивающее взаимодействие информационных систем посредством интернет-протоколов
XML	В данном документе термин используется для обозначения формата файлов обмена данными между информационными системами.

2. Вводная информация

2.1. Назначение документа

Назначением данного документа является описание:

- Краткое описание функциональности ПО «АйТи-Поиск»;
- Краткое описание архитектуры ПО «АйТи-Поиск».

2.2. Описание рассматриваемого ПО

Программное обеспечение «Поисково-аналитическая система АйТи-Поиск» (далее - АйТи-Поиск или ПО АйТи-Поиск) принадлежит компании ООО «Преферентум» (регистрационный №2009611815 в федеральной службе по интеллектуальной собственности (РОСПАТЕНТ), программа для ЭВМ).

ПО АйТи-Поиск реализует функции семантического анализа, извлечения сущностей из документов и проверки документов на соответствие реализованным в системе правилам.

ПО разрабатывается и поддерживается с учетом возможности работы как на платформе MS Windows, так и с использованием только импортозамещающего ПО.

Данный документ содержит описание процессов, обеспечивающих поддержание жизненного цикла программного обеспечения ПО АйТи-Поиск, в том числе устранение неисправностей, выявленных в ходе эксплуатации ПО, совершенствование ПО, а также информацию о персонале, необходимом для обеспечения такой поддержки.

ПО разработано и развивается сотрудниками ООО «Преферентум».

Исключительные права на программное обеспечение АйТи-Поиск принадлежат ООО «Преферентум».

3. Назначение и функциональные возможности ПО «АйТи-Поиск»

ПО «АйТи-Поиск» - поисково-аналитическая система, реализующая функции семантического анализа, извлечения сущностей из документов и проверки документов на соответствие реализованным в системе правилам. ПО «АйТи-Поиск» поставляется в виде прав на использование программного обеспечения (лицензий).

Система обеспечивает следующие функциональные возможности:

- Возможность выделения информационных объектов – организаций, персон, должностных лиц, дат, денежных сумм, валют, условий договора, условия поставки/отгрузки/оплат, названия ТМЦ и Услуг, единиц измерения, нормативной документации (ГОСТ, ТУ), географических объектов, почтовых адресов, ссылок на нормативные правовые акты и пр.;
- Возможность автоматической классификации информации (документов);
- Возможность автоматического формирования в тексте проверяемого документа гипертекстовых ссылок на обнаруженные упоминания нормативных правовых актов и их составных частей в ЛНА;
- Возможность администрирования системы, в том числе возможность добавления новых правил обработки документов;
- Возможность настройки получения и обработки документов из подключаемых источников (например, база ЛНА) и передачу документов в подсистему интеллектуальной обработки информации;
- Возможность сохранения в БД:
 - данных о выделенных фактах и сработавших правилах, в том числе рейтинги срабатывания правил;
 - фрагментов текстов, аннотации и комментарии Экспертов (в объеме, необходимом для настройки имеющихся или создания новых правил);

ПО АйТи-Поиск предоставляет возможность настраивать правила и создавать из них наборы правил (процессы), просматривать рейтинги правил, менять пороговые значения для включения\выключения правила в процессе и просматривать статистику по обработанным документам.

Подробная информация по администрированию системы представлена в

документе «Руководство администратора ПО АйТи-Поиск».

4. Описание работы системы

ПО «АйТи-Поиск» предоставляет возможность с помощью внутренних алгоритмов выделять из неструктурированного текста так называемые «именованные сущности» (информационные объекты), которые позже используются правилами для обработки документа.

Правила являются основным средством системы для обработки документов. Правила кодируются разработчиками на языке C#, загружаются в виде .DLL-файлов и при обработке могут использовать выделенные ранее информационные объекты.

После завершения работы правила, для документа создаются аннотации по отработанным правилам и отправляются во внешнюю систему отображения результатов анализа, а в БД записывается список отработавших правил, выделенные сущности и фрагменты текста с аннотациями.

В системе реализована возможность автоматического формирования в тексте проверяемого документа гипертекстовых ссылок на обнаруженные упоминания нормативных правовых актов и их составных частей в ЛНА. Данную функциональность предоставляет специализированный компонент системы – «Документорий», в который заранее загружаются все необходимые документы.

Компонент системы «Worker» предоставляет механизм обработки документа с помощью правил, а компонент «Manager» запускает заданное администратором число компонентов «Worker»

5. Взаимодействие АйТи-Поиск с внешними ИС

ПО «АйТи-Поиск» предоставляет программный интерфейс (API) для взаимодействия посредством протокола REST в рамках интеграционных взаимодействий.

6. Описание концептуальной архитектуры

В разделе представлена схема концептуальной архитектуры поисково-аналитической системы АйТи-Поиск (см. [Рисунок 1](#)).

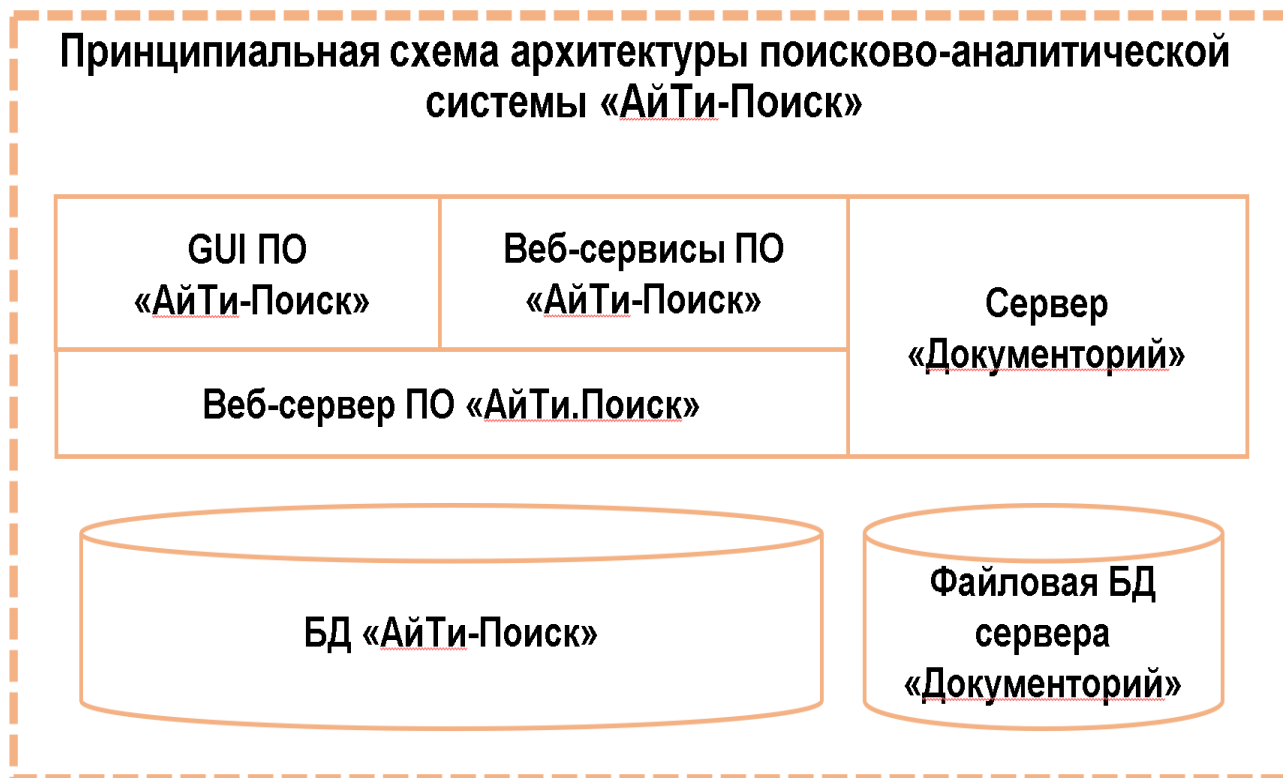


Рисунок 1. Схема концептуальной архитектуры

- «АйТи-Поиск» содержит в себе:
 - БД заданий на обработку и результатов обработки» содержит и связывает между собой задания на обработку документов и результаты их обработки.
 - Интерфейс работы Администратора интеллектуальных правил.
 - Web-сервисы АйТи-Поиск является серверным приложением, выполняющим обработку текстовой информации на основе правил интеллектуальной обработки. Взаимодействует с внешними системами посредством web-сервисов.
 - Правила обработки документов (пакетов документов):
 - наименования правил,
 - их категоризацию и группировку,
 - ссылки на процедуры обработки,
 - и т.д.
 - Фрагменты обработанных документов, аннотации и комментарии Экспертов, - в объеме, необходимом для формирования и настройки интеллектуальных правил Администратором;

- Ссылки на оригиналы ТЗ/ПД и ТКП в БД заданий на обработку и результатов обработки;
- Статистику срабатывания правил;
- Прочую информацию, необходимую для работы правил.
- Модуль обработки ссылок на нормативную документацию (Документорий), состоит из компонент:
 - Сервер «Документорий» реализует функциональность специализированного web-сервиса. Он принимает запросы от внешних приложений на проверку текстовой информации на корректность указания наименований и актуальности срока действия нормативно-правовых документов (далее НПД) и локальных нормативных актов (далее ЛНА), при их наличии в проверяемом тексте, а также позволяет формировать гиперссылки на НПД и ЛНА в аннотациях к проверяемому тексту.
 - «Файловая БД Документория» содержит в себе базу данных Документория, реализованную в виде папки с индексными файлами и файлами с текстами НПД и ЛНА. Папка расположена на том же сервере, где установлен Документорий.
 - «Папка для обновления документов» является «общей папкой», куда Администратор по мере необходимости вносит файлы для пополнения и актуализации Файловой БД Документория. Расположение папки - Сервер «АйТи-Поиск».
- «Классификатор» - сервис, позволяющий по содержимому документа определить (классифицировать) его тип (коммерческое предложение, техническое задание, проект договора и т.п.).
- Описание процессов:
 - «Пользователь» выполняет следующие функции:
 - Ручной запуск процесса обработки документов;
 - Подтверждение/опровержение корректности срабатывания

правила/извлечения сущности;

- Формирование и настройка правил интеллектуальной обработки информации;
- Ввод правил в действие, отключение некорректно работающих правил;
- Обновление файловой БД Документорий;
- АйТи-Поиск сохраняет в БД:
 - Настройки имеющихся правил;
 - Новые правила.
- Web-сервисы АйТи-Поиск проверяет переданный на обработку текст и возвращает:
 - Найденные сущности и аннотации к ним;
 - Аннотации по результатам применения правил обработки.
- При обработке текста компонент Web-сервисы АйТи-Поиск обращается к web-сервису Документория на предмет идентификации фрагментов, являющихся наименованиями НПД и ЛНА. При нахождении таких фрагментов Документорий возвращает информацию об указанном в тексте НПД или ЛНА, включающую корректность наименования документа, актуальность его срока действия, гиперссылку на оригинал документа в Документории.
- Web-сервисы АйТи-Поиск сохраняет в БД:
 - данные о выделенных фактах и сработавших правилах, в том числе рейтинги срабатывания правил;
 - фрагменты текстов, аннотации и комментарии Экспертов (в объеме, необходимом для настройки имеющихся или создания новых правил).