

Министерство образования Республики Беларусь

Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет информационных технологий и управления
Кафедра интеллектуальных информационных технологий

К защите допустить:

Заведующий кафедрой ИИТ

_____ Д. В. Шункевич

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
к курсовому проекту
по дисциплине «Проектирование баз знаний»:

**ПЕРСОНАЛЬНЫЙ АССИСТЕНТ ПО
ЮРИСПРУДЕНЦИИ**

БГУИР КП5 1-40 03 01 044 ПЗ

Студент гр. 121701
Руководитель

Р. В. Липский
Н. В. Зотов

Минск 2023

СОДЕРЖАНИЕ

Перечень условных обозначений	3
Введение	4
1 Анализ подходов к проектированию персонального ассистента по юриспруденции	5
1.1 Описание системы	5
1.1.1 Предметная область юриспруденции	5
1.2 Подходы к разработке баз знаний	6
1.2.1 Продукционные модели	7
1.2.2 Фреймовые модели	8
1.2.3 Формальные логические модели	9
1.2.4 Семантические модели	9
1.3 Средства проектирования персонального ассистента по юриспруденции	10
1.3.1 Технологии построения семантических сетей	10
1.3.2 Языки построения семантических сетей	12
1.4 Аналоги разрабатываемой системы	13
1.4.1 Информационно-поисковая система «ЭТАЛОН-ONLINE»	13
1.4.2 Информационно-правовая система Нормативка.by	13
1.4.3 AI Lawyer	14
1.5 Вывод	14
2 Проектирование персонального ассистента по юриспруденции	16
2.1 Постановка задачи	16
2.2 Описание пользователя персонального ассистента по юриспруденции	16
2.3 Сценарии использования	17
2.4 Архитектура системы	18
2.5 Алгоритмы	21
2.5.1 Алгоритм создания структуры запроса в базе знаний	21
2.5.2 Алгоритм агента нахождения ответа на вопрос Что такое?	21
2.6 Алгоритм агента отправки ответа в чат	21
2.7 Вывод	22
3 Разработка персонального ассистента по юриспруденции	23
3.1 Используемые средства разработки	23
3.1.1 Метасистема OSTIS	23
3.1.2 Docker	23
3.1.3 JESC	24
3.1.4 Kotlin	24
3.1.5 WitAI	24

3.2	Существующая структура базы знаний	25
3.3	Демонстрация функционала системы	25
3.3.1	Функционал для ответов на вопросы на естественном языке	25
3.3.2	База знаний персонального ассистента по юриспруденции	27
3.4	Вывод	30
	Заключение	31
	Список использованных источников	32

ПЕРЕЧЕНЬ УСЛОВНЫХ ОБОЗНАЧЕНИЙ

В курсовом проекте используются следующие условные обозначения:

- ПрО — Предметная область;
- OSTIS — Open Semantic Technology for Intelligent Systems (Открытые Семантические Технологии для Интеллектуальных Систем);
- OWL — Web Ontology Language (Язык Веб Онтологий);
- RDF — Resource Description Framework (Среда Описания Ресурса);
- SC — Semantic Code (Семантический Код);
- SCg — Semantic Code graphical (Семантический Код графический);
- SCn — Semantic Code natural (Семантический Код естественный);
- SCs — Semantic Code string (Семантический Код строковый);
- URI — Uniform Resource Identifier (Унифицированный Идентификатор Ресурса);
- XML — Extensible markup language (Расширяемый Язык Разметки).

ВВЕДЕНИЕ

Юриспруденция - уже одна из наиболее формализованных областей человеческой деятельности, что связано со её спецификой: законы требуют максимальной четкости, понятности и недвусмысленности для устранения необъективного и предвзятого их применения, исключения человеческого фактора.

Тем не менее, юриспруденция охватывает широкий спектр сфер человеческой жизни от семейных взаимоотношений и до международного взаимодействия. Таким образом, юриспруденция - сложная и многогранная наука, а юридические процессы требуют автоматизации.

Ввиду указанных выше особенностей, традиционных автоматизированных систем недостаточно для решения задач в сфере юриспруденции – необходимо создание сложных интеллектуальных систем, интегрирующих самые разные модели решения задач и представления знаний, что включают, но не ограничиваются современными моделями и методами искусственного интеллекта.

Целью данной работы является создание персонального ассистента по вопросам юриспруденции, способного помогать решать юридические задачи различной сложности. Для этого необходимо выполнить следующие задачи:

- Проанализировать предметную область юриспруденции, рассмотреть различные подходы к разработке систем в данной области, существующие системы и решения.

- Спроектировать и разработать базу знаний по юриспруденции, включающую в себя предметную область и онтологию юриспруденции.

- Спроектировать и разработать решатель задач для автоматической синхронизации базы знаний со эталонным банком данных.

В современном мире юриспруденция проникает во все сферы жизни, из чего следует актуальность персонального ассистента по юриспруденции – такая система будет полезна, не только профессионалам, но и простым обывателям, нуждающимся в консультации по юридическим вопросам. Строгая формализация предметной области позволяет обнаружить противоречия и неточности в законах, актах и иных юридических документах, построение онтологии сделает юридическую документацию доступной как для машинной обработки, так и упростит ручное исследование и обработку.

1 АНАЛИЗ ПОДХОДОВ К ПРОЕКТИРОВАНИЮ ПЕРСОНАЛЬНОГО АССИСТЕНТА ПО ЮРИСПРУДЕНЦИИ

1.1 Описание системы

1.1.1 Предметная область юриспруденции

Предметная область юриспруденции – это область знаний, которая изучает право как систему норм и принципов, регулирующих отношения между людьми, организациями и государством. Она включает в себя анализ правовых норм, их источников, принципов интерпретации и применения, правовых институтов и процедур, изучение различных отраслей права.

Кроме того, предметная область юриспруденции включает анализ правовой доктрины, судебной практики и международного права. Изучение юридических систем различных стран и сравнительное правоведение также являются частью предметной области юриспруденции.

В целом, предметная область юриспруденции исследует право как явление, которое регулирует поведение людей и организаций, устанавливает нормы и принципы справедливости, а также определяет процедуры решения юридических споров и конфликтов. Она имеет важное значение для понимания правовой системы, правовых правил и обязанностей, а также для обеспечения справедливости и защиты прав и интересов граждан и организаций.

Таким образом, юридические науки – это часть общественных наук, поскольку государство и право являются социальными институтами. В свою очередь, юридические науки можно подразделить на следующие группы (рис. 1.1):

- теоретико-исторические науки (теория государства и права, история государства и права, история политических и правовых учений
- отраслевые юридические науки (науки конституционного, гражданского, уголовного права и т.п.) [1];
- прикладные науки (криминалистика, криминология, судебная статистика, судебная медицина и т.п.);
- науки международного права;
- организационные науки (судоустройство, организация адвокатуры, нотариат и др.);
- компаративистика (сравнительное правоведение и государствоведение) [2].

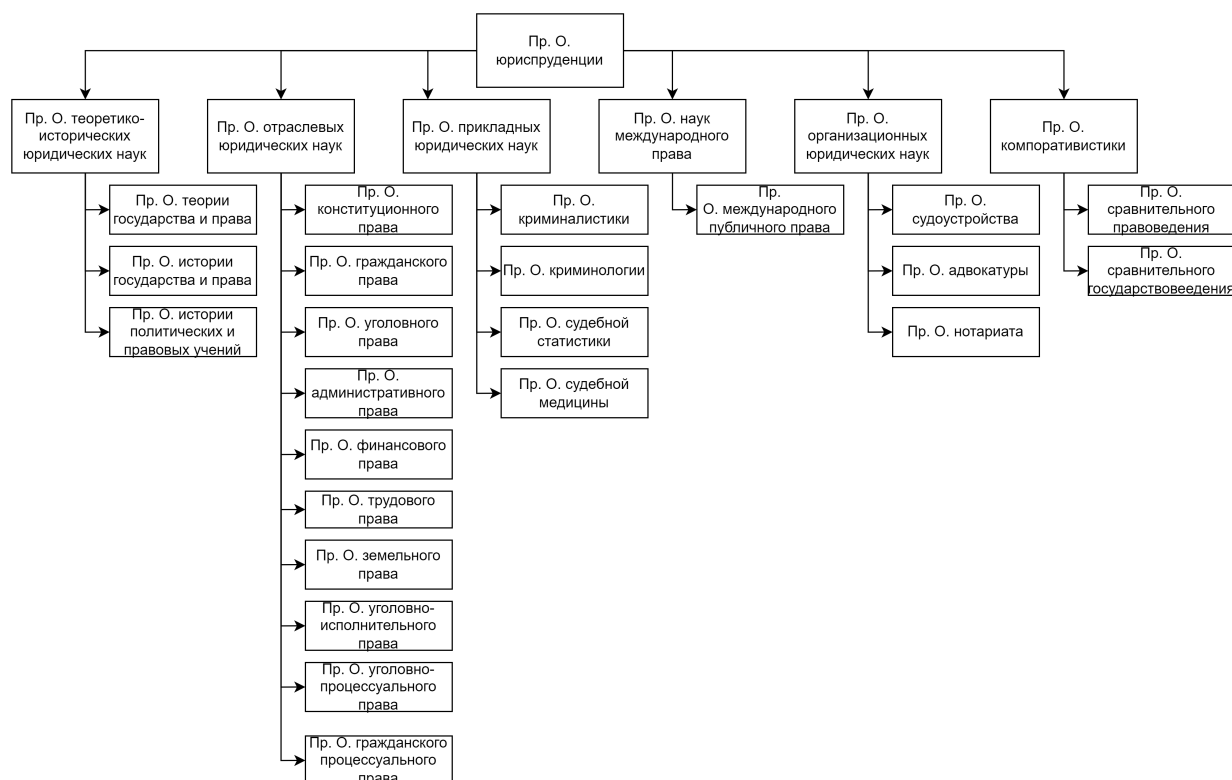


Рисунок 1.1 – Возможная иерархия предметных областей базы знаний персонального ассистента по юриспруденции

1.2 Подходы к разработке баз знаний

Существует важное различие между терминами «данные» и «знания»: данные представляют собой отдельные факты, свойства объектов, процессов и явлений в определенной области, полученные в эмпирическом исследовании. Знания же, в свою очередь, основаны на этих данных, но создаются в результате мыслительной деятельности человека, который обобщает свой опыт, приобретенный в процессе практической работы. Они представляют собой закономерности этой предметной области (принципы, связи, законы), полученные благодаря практическому опыту и профессиональной деятельности специалистов, которые могут применять эти знания для решения конкретных задач. Можно также рассматривать знания как хорошо структурированные данные или как данные о данных (метаданные). [3, с. 19-20]

Существует множество различных моделей представления знаний, которые используются для разных предметных областей, однако большинство из них можно отнести к одним из четырех основных типов:

- продукционные модели;
- фреймовые модели;
- формальные логические модели;
- семантические модели [4, с. 14-15].

1.2.1 Продукционные модели

Продукционная модель или модель, основанная на правилах, позволяет представить знания в виде предложений, типа «если (*условие*), то (*действие*)».

Под «условием» (*антицедентом*) понимается некоторое предложение-образец, по которому осуществляется поиск в базе знаний, а под «действием» (*консеквентом*) – действия, выполняемые в успешном исходе поиска (они могут быть промежуточными, выступающими далее как условия и терминальными или целевыми, завершающими работу системы).

Чаще всего вывод на такой *базе знаний* бывает **прямой** (от данных к поиску цели) или **обратный** (от цели её подтверждения – к данным). *Данные* в этом контексте – исходные факты, хранящиеся в *базе фактов*, на основании которых запускается *машина вывода* или *интерпретатор правил*, перебирающий правила из *продукционной базы знаний*. [3, с. 21-22]

Системы produkcji с прямыми выводами являются одним из основных типов систем, использующих знания. Они включают три компонента: *базу правил*, *базу данных* и *интерпретатор для логического вывода*. В процессе выполнения цикла «понимание – выполнение» база данных обновляется на основании выбранных правил, пока не будет достигнут целевой результат. Однако, с увеличением числа правил, скорость вывода замедляется, что ограничивает использование этих систем для крупномасштабных задач.

Система produkcji с обратными выводами использует правила для построения дерева, которое связывает факты и заключения. Оценка этого дерева основывается на фактах в базе данных, и результатом является логический вывод. Обратные выводы имеют преимущество в том, что они оценивают только те части дерева, которые имеют отношение к заключению. Однако, если отрицание или утверждение невозможны, то порождение дерева лишено смысла. В *двунаправленных выводах* сначала оценивается небольшой объем полученных данных и выбирается гипотеза, а затем запрашиваются данные, необходимые для принятия решения о пригодности данной гипотезы. Это позволяет создать более мощную и гибкую систему.

Сильные стороны продукционной модели:

- простота создания и понимания отдельных правил;
- простота модификаций и расширения базы знаний;
- простота механизма логического вывода;
- естественная модульность организации баз знаний.

Слабые стороны продукционной модели:

- сложность оценки целостного образа знаний;
- крайне низкая эффективность обработки;
- отличие от человеческой структуры знаний;

– отсутствие гибкости в логическом выводе [4, с. 15-16].

1.2.2 Фреймовые модели

Фрейм – это абстрактный образ для представления некоего стереотипа восприятия. Термин *фрейм* был предложен Марвином Минским, одним из пионеров искусственного интеллекта для обозначения структуры знаний для восприятия пространственных сцен. Эта модель, как и семантическая сеть, имеет глубокое психологическое обоснование.

Например, произнесение слова, такого как «комната», вызывает у слушателей определенный образ в их воображении. Этот образ включает в себя основные характеристики комнаты, такие как четыре стены, пол, потолок, окна и дверь. Нельзя убрать эти основные характеристики, такие как окна, иначе мы перестаем говорить о комнате и начинаем говорить о другом объекте, например, чулане. Но этот абстрактный образ также имеет «*слоты*» - неопределенные характеристики, такие как количество окон, цвет стен, высота потолка и т.д.

В теории фреймов такой образ комнаты называется фреймом комнаты. Фреймом называется также и формализованная модель для отображения образа. Различают *фреймы-образцы (прототипы)*, хранящиеся в базе знаний и *фреймы-экземпляры*, которые создаются для отображения реальных фактических ситуаций на основе поступающих данных. Модель фрейма является достаточно универсальной, поскольку позволяет отобразить всё многообразие знаний о мире через:

- фреймы-*структуры*, использующиеся для обозначения объектов и понятий (заём, залог...);
- фреймы-*роли* (менеджер, кассир, клиент...);
- фреймы-*сценарии* (банкротство, собрание, празднование...);
- фреймы-*ситуации* (авария, тревога...).

Теория фреймов заимствует важное свойство из теории семантических сетей, которое называется "заимствование свойств". В обеих теориях наследование осуществляется через АКО-связи (A-Kind-Of = это), которые указывают на фрейм или понятие более высокого уровня в иерархии. Благодаря этому наследованию свойства передаются от более общего к более конкретному фрейму. [3, с. 23-25]

Основным преимуществом фреймов как модели представления знаний является то, что она отражает концептуальную основу памяти человека, а также её гибкость и наглядность. [3, с. 25]. Также фреймы обеспечивают требования структурированности и связанности. Это достигается за счёт свойств наследования и вложенности, которыми обладают фреймы. [4, с. 17]

1.2.3 Формальные логические модели

Логическая модель используется для представления знаний в системе логики предикатов первого порядка и логики высказываний.

Формальная дедуктивная система, лежащая в основе логической модели представления знаний, может быть представлена в виде четверки $f = \langle V, S, A, P \rangle$, где V – множество базовых элементов (алфавит); S – множество синтаксических правил, на основе которых из V строятся правильно построенные формулы; A – множество аксиом (правильно построенных формул); P – множество правил вывода (семантических правил), которые из множества аксиом позволяют получать новые правильно построенные формулы – теоремы.

Использование логики предикатов для представления знаний имеет ряд преимуществ. Во-первых, такой подход обладает четкими математическими свойствами и позволяет создавать мощные механизмы вывода, которые можно непосредственно запрограммировать. Во-вторых, такой подход обеспечивает контроль логической целостности базы знаний, что позволяет гарантировать ее непротиворечивость и полноту. В-третьих, запись фактов в такой модели очень проста и компактна.

Однако главным недостатком логической модели является отсутствие четких принципов организации фактов в базе знаний. Из-за этого обработка и анализ больших баз знаний может быть затруднительным. [4, с. 15]

1.2.4 Семантические модели

Семантическая сеть – это орграф, в котором вершины представляют понятия, а дуги отображают отношения между ними. Термин *семантическая* означает «относящийся к смыслу», «смысловая», а «семантика» – это наука, которая определяет отношения между символами и объектами, которые они обозначают, то есть наука, определяющая смысл знаков. В качестве понятий в семантической сети обычно выступают абстрактные или конкретные объекты, а отношения – это связи между этими объектами, типа: «это» («АКО – A Kind Of», «is»), «имеет частью» («has part»), «принадлежит», «любит» и т.д. Характерной особенностью семантических сетей является обязательное наличие трёх типов отношений: «**класс** – элемент класса», «**свойство** – значение», «**пример** элемента класса.».

Можно предположить несколько классификаций семантических сетей, связанных с типами отношений между понятиями: по количеству различных типов отношений (однородные – с единственным типом отношений, неоднородные – с различными типами отношений), по самим типам этих отношений (бинарные – отношения связывают ровно два объекта, N-арные – отношения могут связывать два и более объектов).

Проблема поиска решения в базе знаний типа семантической сети сводится к задаче поиска фрагмента сети, соответствующего некоторой подсети, отражающей поставленный запрос к базе. [3, с. 22-23]

Достоинствами моделей представления знаний с помощью семантических сетей являются большие выразительные возможности, наглядность, близость структуры сети семантической структуре предметной области [4, с. 18], соответствие современным представлениям об организации долговременной памяти человека. [3, с. 23]

Семантические сети - это метод представления знаний, который имеет недостатки в теоретическом развитии логического вывода, что приводит к увеличению произвольности, введенной разработчиком. Поэтому процедуры вывода в таких системах могут быть противоречивыми, и необходимо уделять больше внимания анализу противоречивости. Эту функцию в семантических моделях зачастую выполняет человек, что усложняет решение задач при большом объеме знаний и ограничивает области применения. [4, с. 18] Также недостатком семантической модели является сложность процедуры поиска вывода. [3, с. 23]

1.3 Средства проектирования персонального ассистента по юриспруденции

1.3.1 Технологии построения семантических сетей

Semantic web – это технология, которая расширяет возможности существующего веба, объединяя данные из разных источников. Информация, которая доступна в Интернете, может быть прочитана человеком, но Semantic Web создан для того, чтобы сделать ее пригодной для автоматического анализа и синтеза выводов. Он использует способ представления данных в виде семантической сети с помощью онтологий, что отличается от классического веба. Семантическая паутина использует языки описания, такие как XML, XML Schema, RDF, RDF Schema, OWL и другие, для технической реализации. Она позволяет преобразовывать данные и выводы, основанные на этих данных, в различные представления, которые могут быть полезны на практике. [5]

Преимущества Semantic Web:

- высокая популярность и распространенность стандарта;
- простота в изучении и использовании.

Наиболее обсуждаемыми проблемами инструментария Semantic web являются:

- неоднозначность при определении классов и их экземпляров;
- отсутствие возможности определять свойства у свойств, что не позволяет моделировать атрибуты у предметных отношений, n-арные отно-

шения и атрибуты у атрибутов;

- ориентация на web и близкое к машинному представление семантических сетей;

- неразвитые стандарты представления переменных во времени и нечетких предметных областей;

- слабый уровень верификации онтологий на противоречивость и полноту. [6]

OSTIS – это технология, основанная на семантических сетях, которая позволяет представлять знания. Она предназначена для компонентного проектирования систем, которые управляют знаниями. Предлагаемая технология оформляется как интеллектуальная метасистема, которая строится на основе предлагаемой технологии и содержит в себе все модели, средства и методы, связанные с данной технологией.

Информация представляется в виде логико-семантической модели, которая является платформенно-независимой и позволяет интеграцию с логико-семантическими моделями других систем. Она полностью отображает семантику используемых знаний и методов решения задач.

В процессе проектирования систем, управляемых знаниями, необходимо обращать особое внимание на технологию обновления систем в процессе их эксплуатации, а также на метатеchnологию постоянного совершенствования самой технологии компонентного проектирования. Результатом проектирования логико-семантической модели любой интеллектуальной системы является текст, содержащий исходный текст базы знаний, исходные тексты программ и документацию. [7]

Преимущества технологии OSTIS:

- имеет строгую теоретико-множественную трактовку и не привязаны к конкретной прикладной области;

- имеет возможность представления отношений любой арности;

- отношения представляются в виде узлов семантической сети, что позволяет характеризовать их свойства;

- экземпляры отношений выделяются как отдельные узлы семантической сети, что дает возможность характеризовать каждый экземпляр отношения уникальным образом;

- в алфавите ключевых узлов и дуг имеются элементы для описания нечетких, негативных и нестационарных объектов. [6]

Недостатки технологии OSTIS — это малое количество доступной документации, а также малое количество обучающих материалов с низким порогом вхождения, низкая популярность и распространенность стандарта в мире, высокая ресурсоемкость и низкая стабильность работы ostis-систем на данный момент.

1.3.2 Языки построения семантических сетей

OWL является языком для создания и определения веб-онтологий. Он описывает классы, свойства и их экземпляры, а также отношения между ними. OWL используется для явного представления смысла терминов и словарных отношений. Язык OWL предоставляет три подязыка, каждый из которых предназначен для конкретных сообществ исполнителей и пользователей:

- OWL Lite – наименее выразительный, подходит для простых иерархий классов и ограничений, удобен для миграции тезаурусов и таксономий.
- OWL DL – более выразительный, сохраняет вычислительную полноту, гарантирует вычислимость выводов и основан на логическом описании.
- OWL Full – наиболее выразительный, имеет синтаксическую свободу RDF, не обеспечивает вычислительные гарантии, но позволяет расширять смысл заранее определенных словарей. [8]

OWL является развитием технологий XML и RDF и входит в новый стек веб-протоколов. В отличие от многих языков представления знаний, OWL не является исключительно академическим, а представляет практическую ценность. Он получает финансовую поддержку и имеет множество программных инструментов для работы.

С точки зрения полноты выразительных возможностей язык OWL не может конкурировать со многими существующими языками представления знаний уже хотя бы потому, что не рассчитан на описание динамических явлений (его основа – это объекты и связи). [9]

SC-код – используется для организации сложной структурированной информации, путем унификации различных видов знаний в общую форму. Этот принцип позволяет интегрировать знания в базе знаний и объединять их. Существуют различные внешние представления SC-кода, такие как SCg, который представляет содержимое базы знаний в виде графа; SCn, приближенный к естественному языку; SCs, который представляет содержимое базы знаний в виде строки.

Достоинствами SC-кода являются следующие его свойства:

- Неограниченная возможность перехода от sc-текстов к sc-метатекстам, содержащим знаки описываемых sc-текстов;
- Тексты SC-кода могут быть иерархическими структурами, поскольку sc-элемент может обозначать множество, состоящее из любых sc-элементов;
- Все основные семантические связи между текстами в SC-коде становятся теоретико-множественными.

SC-код объединяет в себе язык и метаязык и может использоваться для описания синтаксиса, семантики и онтологии этого кода. Он работает только с семантически нормализованными множествами, что позволяет

четко разделять первичные (терминальные) элементы, которые являются обозначениями внешних объектов, и вторичные элементы, которые являются обозначениями множеств и могут быть использованы в составе SC-конструкций. После формирования синтаксиса строится модель, которая может описывать фрагмент какой-либо предметной области. [10, с 116-117]

1.4 Аналогии разрабатываемой системы

На данный момент уже существуют некоторые информационные системы, предоставляющие пользователям услуги в сфере юриспруденции.

1.4.1 Информационно-поисковая система «ЭТАЛОН-ONLINE»

Информационно-поисковая система «ЭТАЛОН-ONLINE», разработанная Национальным центром правовой информации. ЭТАЛОН-ONLINE предоставляет доступ к правовой информации Республики Беларусь, а именно

- банк данных (БД) «Законодательство Республики Беларусь»;
- БД «Решения органов местного управления и самоуправления»;
- БД «Международные договоры»;
- БД «Судебная практика»;
- БД «Правоприменительная практика»;
- БД «Формы документов». [11]

Также данная система предоставляет доступ к набору инструментов, таких как юридический словарь, калькуляторы (например, калькулятор трудового стажа, калькулятор трудовых дней, подлежащих компенсации и т.д.), конструкторы документов, проверка контрагентов и т.д.

Система не является интеллектуальной и не претендует на это – она не способна решать прикладные задачи, за исключением некоторых подсчётов, не содержит формализованных знаний о предметной области и даже не имеет программного интерфейса.

Тем не менее, система может оказаться полезна для разработчиков других систем, поскольку может служить источником данных для построения формальных онтологий, обучения каких-либо моделей и т.д.

1.4.2 Информационно-правовая система Нормативка.by

Информационно-правовая система Нормативка.by, которая обеспечивает бухгалтеров, юристов, работников кадровых служб, экономистов полным комплексом правовой, аналитической и справочной информации [12].

Система предоставляет официальную правовую информацию, которая поступает из Эталонного банка правовой информации Республики Беларусь, предоставляет возможность быстрого и эффективного поиска по содержимой информации, а также предоставляет пошаговые алгоритмы действий по наиболее актуальным и сложным вопросам.

По сути своей, данная система – просто альтернативный интерфейс к ЭТАЛОН-ONLINE и не предлагает ничего большего, чем может предложить ЭТАЛОН-ONLINE.

1.4.3 AI Lawyer

AI Lawyer – система, предоставляющая функции персонального ассистента по юриспруденции. Разработчики заявляют, что система способна заменить юристов и предоставлять юридическую справку и советы в реальном времени. [13]

В действительности, AI Lawyer является лишь модификацией чат-бота ChatGPT, разработанной компанией OpenAI. И хотя ChatGPT с моделью GPT-4 показывает хорошие результаты в поддержании диалога на различные темы, чат-бот склонен допускать простейшие ошибки, может использовать устаревшую, ложную и просто выдуманную информацию для обоснования своих выводов.

Так, если задать вопрос, связанных с законодательством, система прямо отвечает, что «не является юристом и может предоставлять лишь общую информацию о законодательстве».

Существует и другие аналогичные системы, использующие ChatGPT для предоставления юридических услуг, такие как Legal ChatGPT, LegalAI и т.д., которые разделяют все недостатки AI Lawyer.

1.5 Вывод

На данный момент не существует единой публично доступной базы знаний, полностью охватывающей всю предметную область юриспруденции. Хотя существуют различные ресурсы и базы данных, содержащие юридическую информацию, они обычно ограничены определенными юрисдикциями, отраслями права или типами юридических материалов.

Существующие решения, такие как ЭТАЛОН-ONLINE, Нормативка.by, не предоставляют широких возможностей для использования пользователем без экспертизы в сфере юриспруденции, поскольку и не ставят это своей целью. Также необходимо заметить, что решать хоть сколько-нибудь нетривиальные задачи при помощи этих систем невозможно, ввиду отсутствия каких-либо инструментов, позволяющих это делать.

Система «Персональный ассистент по юриспруденции» нацелена на

то, чтобы позволить обычному пользователю, не обладающему профессиональными знаниями получать актуальную и достоверную юридическую информацию, решать различные прикладные задачи с представлением обоснования решения.

В связи с преимуществами семантической модели представления знаний, для построения фрагмента базы знаний было решено использовать технологию OSTIS. Из языков внешнего представления SC-кода был выбран SCs (Semantic Code string), поскольку он наиболее удобен для совместной разработки с использованием Git.

2 ПРОЕКТИРОВАНИЕ ПЕРСОНАЛЬНОГО АССИСТЕНТА ПО ЮРИСПРУДЕНЦИИ

2.1 Постановка задачи

– Разработать базу знаний персонального ассистента по юриспруденции, которая будет включать в себя предметную область и онтологию различных областей юриспруденции, формальное описание понятий и ссылки на законодательные акты, определяющие эти понятия. В частности, сделать упор на формализацию белорусского законодательства.

– Разработать решатель задач, позволяющий задавать вопросы к системе в свободной форме на естественном языке и получать ответы.

2.2 Описание пользователя персонального ассистента по юриспруденции

Персональный юридический ассистент может быть полезен различным типам пользователей, включая:

– Частных лиц, которые нуждаются в справочной информации и пояснениях о юриспруденции для своих личных нужд. Например, они могут обратиться к ассистенту для получения информации о правовых процедурах, правах и обязанностях граждан, процессе регистрации брака или развода, праве на наследство и других вопросах, связанных с их личной жизнью.

– Студентов юридических специальностей, которые могут использовать ассистента для получения дополнительной информации и пояснений по учебным предметам. Они могут обратиться к ассистенту для изучения различных областей юриспруденции, понимания правовых терминов, изучения прецедентов и других материалов, которые могут помочь им в учебе.

– Малого бизнеса и предпринимателей: Владельцы малых бизнесов и предприниматели могут использовать ассистента для получения справочной информации по вопросам, связанным с правовыми аспектами бизнеса. Например, они могут обратиться к ассистенту для получения информации о процессе регистрации компании, налоговых обязательствах, правилах ведения бизнеса, правах и обязанностях работников и других юридических вопросах, связанных с их предпринимательской деятельностью.

– Люди, которые сталкиваются с конкретными юридическими вопросами или проблемами, могут обратиться к ассистенту для получения справочной информации и советов. Например, они могут обратиться к ассистенту для получения информации о процессе подачи искового заявления, правилах и сроках исковой давности, процедурах разрешения споров и других вопросах, связанных с их конкретной ситуацией.

Каждому из этих типов пользователей предоставляются соответствующие

ющие функциональности и сценарии использования, чтобы обеспечить максимально эффективное использование персонального юридического ассистента и удовлетворение их индивидуальных потребностей в справочной информации и пояснениях о юриспруденции.

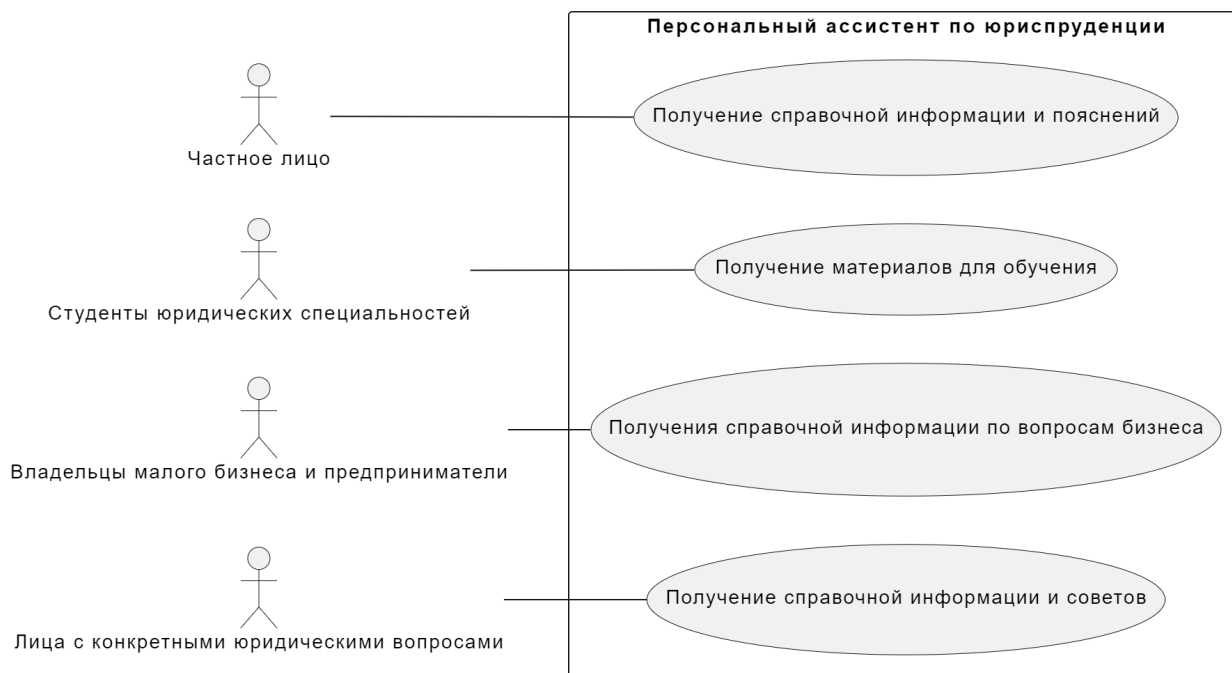


Рисунок 2.1 – Сценарии использования персонального ассистента по юриспруденции

2.3 Сценарии использования

Сценарий 1

- Пользователь открывает Telegram и находит чат с OSTIS Legislation Bot.
- Пользователь пишет в чат: "Что такое исковая усыновление?"
- OSTIS Legislation Bot принимает сообщение и создает запрос в своей базе знаний, ища информацию о понятии "усыновление".
- Агент нахождения ответа анализирует запрос и находит соответствующую информацию о понятии "усыновление" в базе знаний.
- Агент отправления ответа передает найденное определение и объяснение понятия обратно в чат Telegram.
- Пользователь получает ответ от OSTIS Legislation Bot, содержащий определение и объяснение понятия "усыновление".

Сценарий 2

- Пользователь открывает Telegram и находит чат с OSTIS Legislation Bot.
- Пользователь пишет в чат: "В каком документе дано определение понятию 'договор'?"

- OSTIS Legislation Bot принимает сообщение и формирует запрос, ища информацию о документе, в котором содержится определение понятия "договор".

- Агент нахождения ответа анализирует запрос и находит соответствующую информацию о документе, содержащем определение понятия "договор" в базе знаний.

- Агент отправления ответа передает найденную информацию о документе обратно в чат Telegram.

- Пользователь получает ответ от OSTIS Legislation Bot, содержащий информацию о документе, в котором дано определение понятия "договор".

Сценарий 3

- Пользователь открывает Telegram и находит чат с OSTIS Legislation Bot.

- Пользователь пишет в чат: "О чем говорится в статье 15 Кодекса о браке и семье Республики Беларусь"

- OSTIS Legislation Bot принимает сообщение и создает запрос, ища информацию о содержании статьи 15 Кодекса о браке и семье Республики Беларусь.

- Агент нахождения ответа анализирует запрос и находит соответствующую информацию о содержании статьи 15 Кодекса о браке и семье Республики Беларусь в базе знаний.

- Агент отправления ответа передает найденную информацию о содержании статьи 15 Кодекса обратно в чат Telegram.

- Пользователь получает ответ от OSTIS Legislation Bot, содержащий информацию о том, о чем говорится в статье 15 Кодекса о браке и семье Республики Беларусь.

2.4 Архитектура системы

Система «Персональный ассистент по юриспруденции» должна иметь следующую архитектуру:

- База знаний персонального ассистента по юриспруденции
- Решатели задач персонального ассистента по юриспруденции
- Интерфейс персонального ассистента по юриспруденции

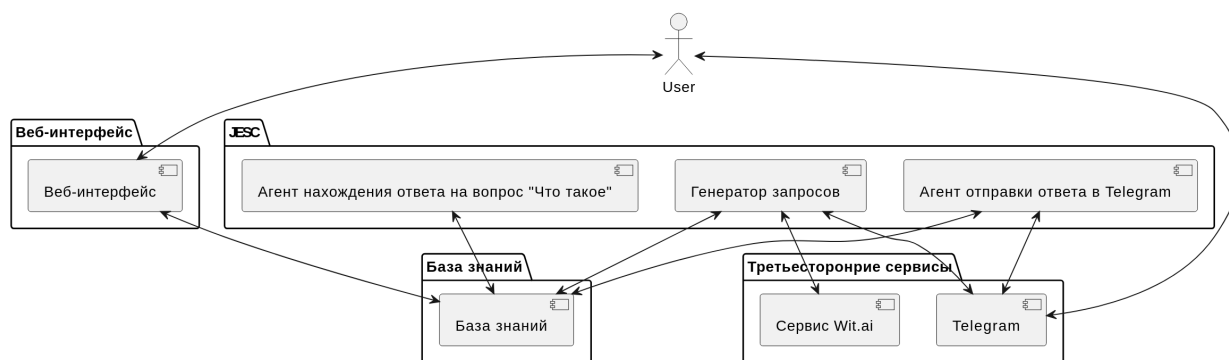


Рисунок 2.2 – Диаграмма компонентов персонального ассистента по юриспруденции

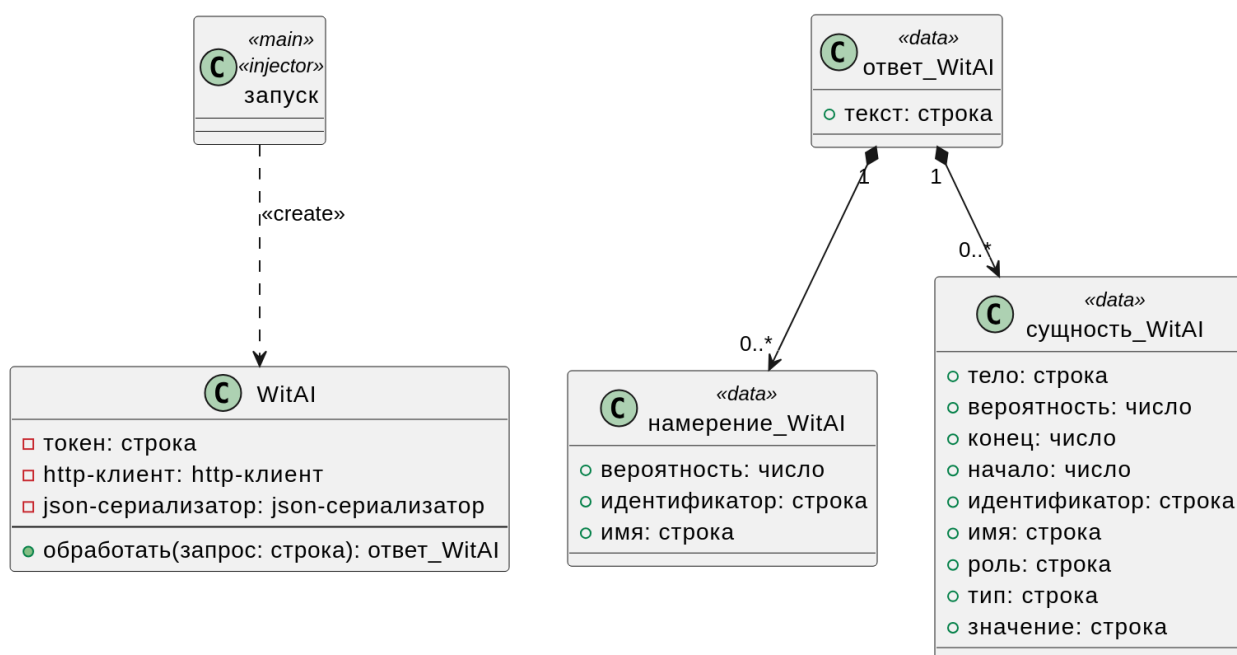


Рисунок 2.3 – Диаграмма классов для клиента WitAI

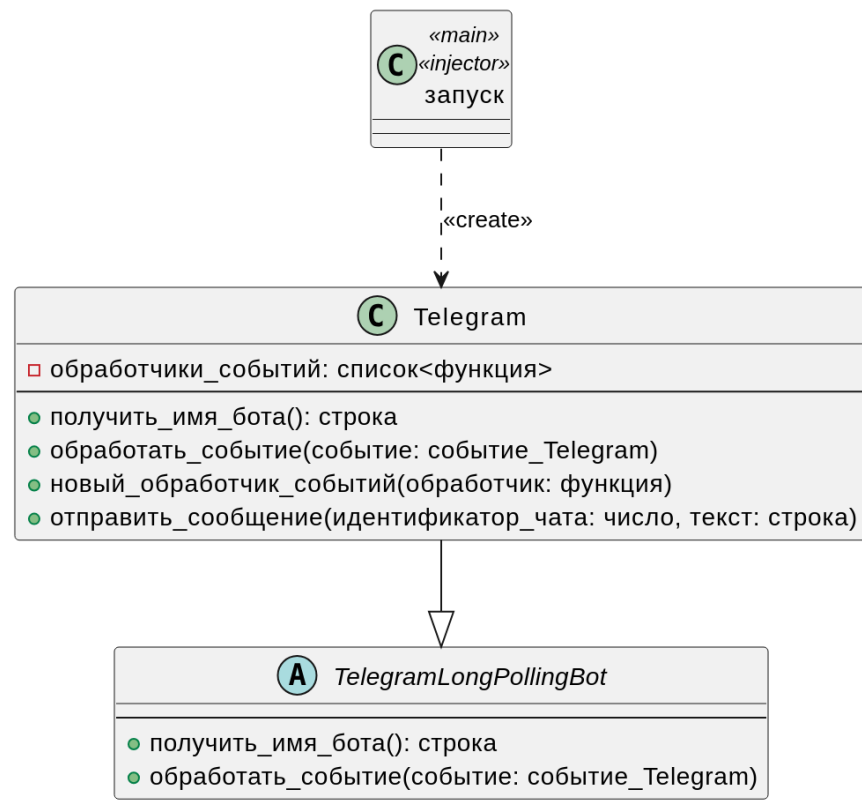


Рисунок 2.4 – Диаграмма классов для Telegram-клиента

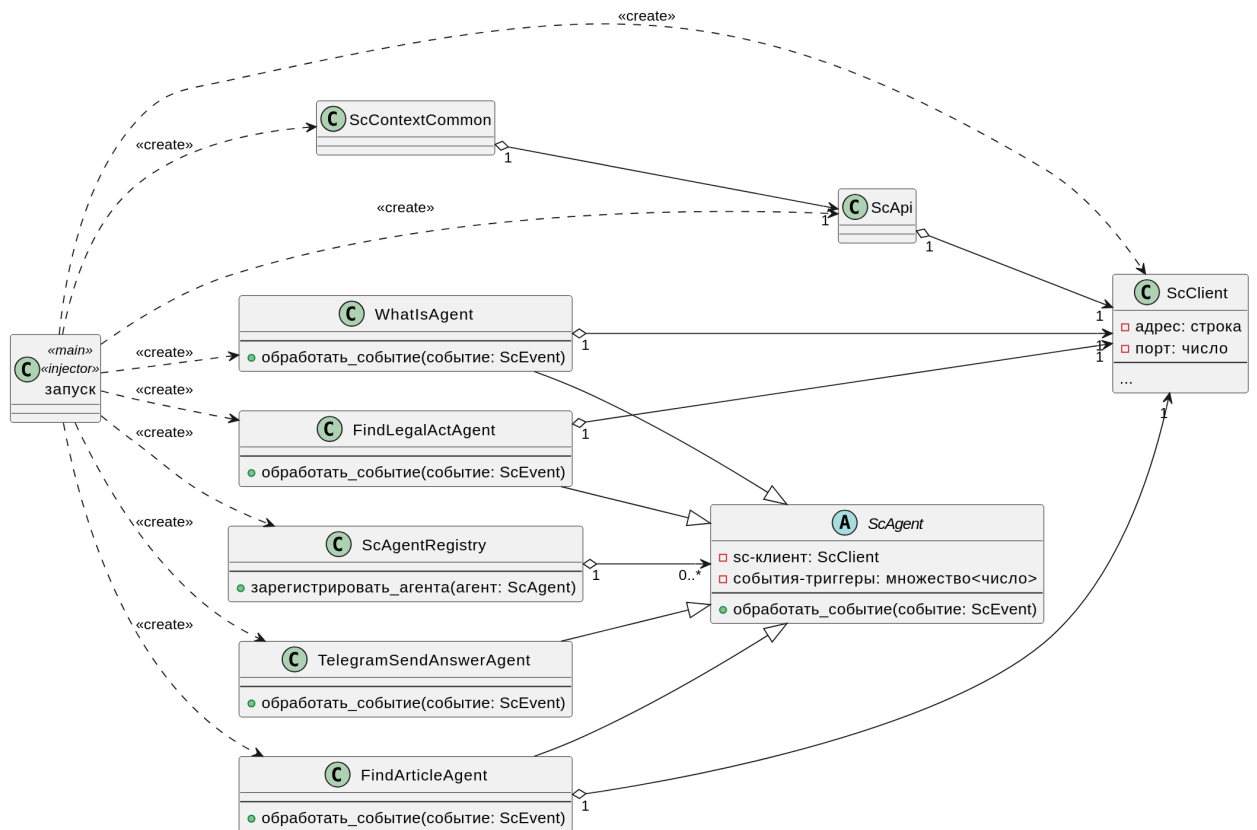


Рисунок 2.5 – Диаграмма классов для решателей задач

2.5 Алгоритмы

2.5.1 Алгоритм создания структуры запроса в базе знаний

- Получить результат обработки запроса на естественном языке от Wit.AI.
- Создать узел, являющийся родительским для всей структуры;
- Создать файл, содержащий название искомого понятия;
- Создать файл, содержащий идентификатор чата;
- Создать дугу принадлежности (1) от структуры к сущности, определяющей намерение пользователя
- Создать дугу принадлежности (2) от структуры к файлу с названием искомого понятия
- Создать дугу принадлежности (3) от структуры к файлу с идентификатор чата
- Создать дугу принадлежности от множества отношений "намерение" к дуге (3).
- Создать дугу принадлежности от множества отношений "первый аргумент" к дуге (4).
- Создать дугу принадлежности от множества отношений "идентификатор чата" к дуге (5).

2.5.2 Алгоритм агента нахождения ответа на вопрос Что такое?

- Если намерение пользователя не равно Что такое?, выполнение завершается.
- Получить название искомого понятия из базы знаний.
- Получить названия всех понятий, содержащихся в базе знаний.
- Найти наиболее близкое по названию понятие.
- Получить определение искомого понятия.
- Создать дугу принадлежности от структуры запроса к файлу с определением.
- Создать дугу принадлежности от множества отношений "ответ" к предыдущей дуге.

2.6 Алгоритм агента отправки ответа в чат

- Найти файл, связанный отношением "ответ" с структурой запроса.
- Найти файл, связанный отношением "идентификатор чата" со структурой запроса.
- Отправить содержимое файла в чат с пользователем.

2.7 Вывод

Результатом проектирования стал проект системы «Персональный ассистент по юриспруденции» – были уточнены задачи, решаемые системой, были описаны потенциальные пользователи системы, рассмотрены вопросы и задачи, которые пользователь может решать при помощи персонального ассистента по юриспруденции, описаны сценарии использования системы пользователем, архитектура системы, структурированы требования к реализации решателей задач.

3 РАЗРАБОТКА ПЕРСОНАЛЬНОГО АССИСТЕНТА ПО ЮРИСПРУДЕНЦИИ

3.1 Используемые средства разработки

3.1.1 Метасистема OSTIS

Метасистема OSTIS - это универсальная базовая (предметно-независимая) система автоматизации проектирования ostis-систем.

В основе OSTIS лежит:

- формальная онтология интеллектуальных систем, а точнее, формальная онтология их логико-семантических моделей;
- формальная онтология проектирования интеллектуальных систем – формальная онтология средств и методов проектирования логико-семантических моделей интеллектуальных систем, а также проектных действий, направленных на разработку указанных логико-семантических моделей.

Компьютерные системы, разработанные по Технологии OSTIS, можно считать новым поколением компьютерных систем, которые, независимо от решаемых ими задач, становятся обладателями таких свойств, как:

- совместимость;
- платформенная независимость;
- гибкость;
- адаптивность;
- обучаемость [10].

3.1.2 Docker

Docker — это открытая платформа для разработки, доставки и запуска приложений. Docker предоставляет возможность упаковать и запустить приложение в слабо изолированной среде, называемой контейнером [14].

Преимуществами Docker являются:

- Платформенная независимость – приложение, контейнеризированное при помощи Docker возможно развернуть и запустить на всех наиболее популярных платформах.
- Контейнеры легковесны и содержат в себе все необходимое для запуска приложения: операционную систему, программное обеспечение, библиотеки, конфигурацию.
- Удобство распространения и совместного использования контейнеров.

3.1.3 JESC

JESC – библиотека, написанная на языке Java для обеспечения взаимодействия с SC-машиной при помощи её программного интерфейса [15].

Преимуществами JESC являются:

- Удобство в использовании, наличие документации.
- Библиотека поддерживается и обновляется в соответствии с обновлениями SC-машины.
- Библиотека поддерживает различные языки программирования, такие как Java и Kotlin.

3.1.4 Kotlin

Kotlin — безопасный, совместимый с Java и другими языками язык программирования, который предоставляет способов повторного использования кода на нескольких платформах для продуктивного программирования [16].

Преимуществами Kotlin являются:

- Язык предоставляет богатый набор функций и возможностей языка, поддерживает лаконичный синтаксис, расширения функций, функциональное программирование.
- Это язык со статической типизацией, который призван предотвратить ошибки типизации на этапе компиляции. Он предлагает механизмы, такие как инференция типов, нуллабельные и неизменяемые типы данных, безопасные приведения типов и другие возможности, которые способствуют созданию более надежных и безопасных программ.
- Kotlin полностью совместим с Java, что позволяет использовать существующий Java-код и библиотеки в проектах на Kotlin.

3.1.5 WitAI

WitAI - это сервис, разработанный для обработки естественного языка. Он предоставляет разработчикам API и инструменты для создания различных приложений, которые могут взаимодействовать с пользователем через текстовые или голосовые команды.

Главной особенностью Wit.AI является его способность понимать и интерпретировать намерения и контекст ввода пользователя на естественном языке. Он использует машинное обучение и алгоритмы обработки естественного языка для анализа текстовых вводов и извлечения значимых сущностей, намерений и контекста [17].

3.2 Существующая структура базы знаний



Рисунок 3.1 – Существующая структура базы знаний

3.3 Демонстрация функционала системы

3.3.1 Функционал для ответов на вопросы на естественном языке

Ответы на вопросы класса «Что такое?» – система осуществляет поиск по базе знаний и возвращает пользователю определение, соответствующее искомому понятию:

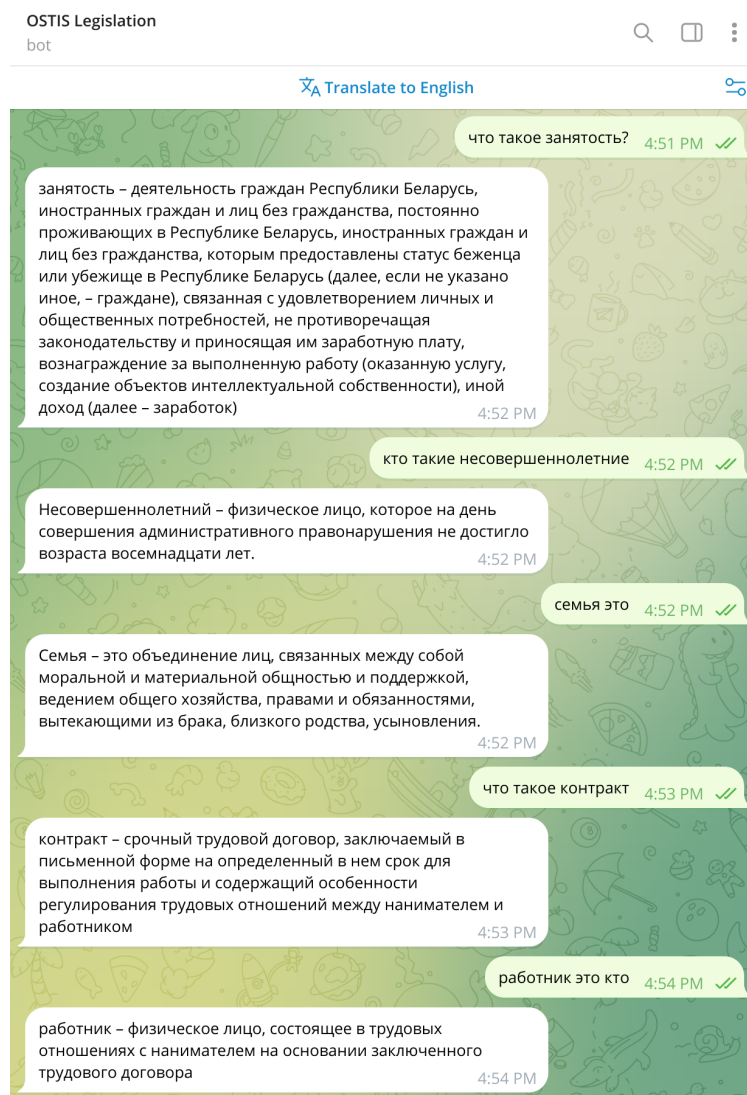


Рисунок 3.2 – Примеры ответов на вопросы "Что такое?"

Ответы на вопросы класса «В каком документе дано определение понятию?» – система осуществляет поиск по базе знаний и возвращает пользователю документ (кодекс, закон), в котором дано определение искомому ПОНЯТИЮ:



Рисунок 3.3 – Примеры ответов на вопросы класса "В каком документе дано определение понятию?"

Ответы на вопросы класса "О чем говорится в статье? система осуществляет поиск по базе знаний и возвращает пользователю содержимое статьи:

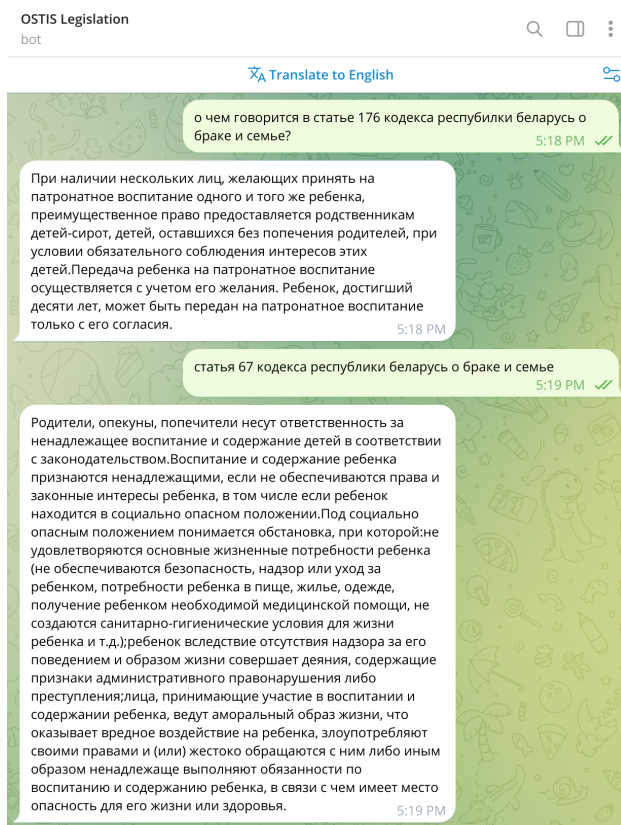


Рисунок 3.4 – Примеры ответов на вопросы класса "О чем говорится в статье?"

3.3.2 База знаний персонального ассистента по юриспруденции

Стартовая страница базы знаний персонального ассистента по юриспруденции – «База знаний по юриспруденции», содержащая разделы «Юриспруденция», «Законодательство Республики Беларусь» (рис. 3.5).

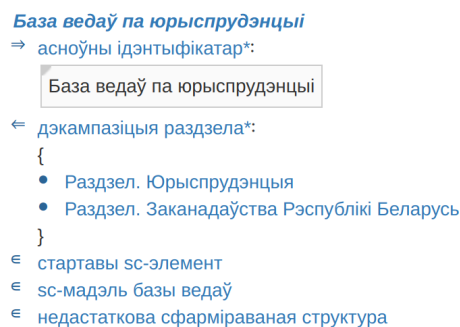


Рисунок 3.5 – Стартовая страница базы знаний персонального ассистента по юриспруденции

Раздел «Белорусское законодательство» содержит в себе предметную область белорусского законодательства, а также разделы, соответствующие различным областям законодательства Республики Беларусь, такие, как «Технические нормативно-правовые акты Республики Беларусь», «Законодательство о браке и семье Республики Беларусь», «Законодательство о труде и занятости населения Республики Беларусь», «Жилищное законодательство Республики Беларусь».

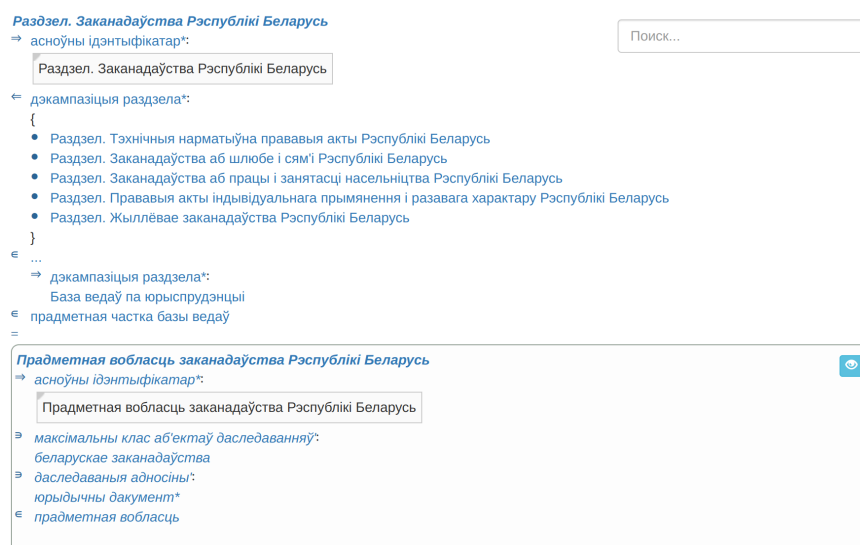


Рисунок 3.6 – Раздел. Белорусское законодательство

В разделах, соответствующих областям белорусского законодательства содержится соответствующая предметная область, а также подразделы – словарь по законодательству, а также раздел, содержащий законодательные акты, относящиеся к данной области законодательства (рис. 3.7).

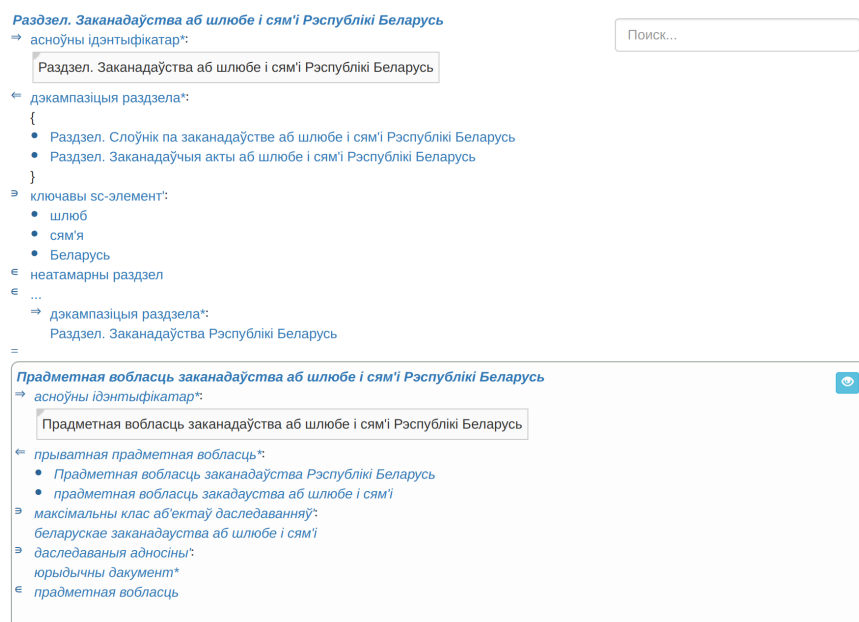


Рисунок 3.7 – Раздел. Законодательство о браке и семье Республики Беларусь

В разделе словаря по области законодательства содержатся понятия, связанные с данной областью законодательства (рис. 3.8).

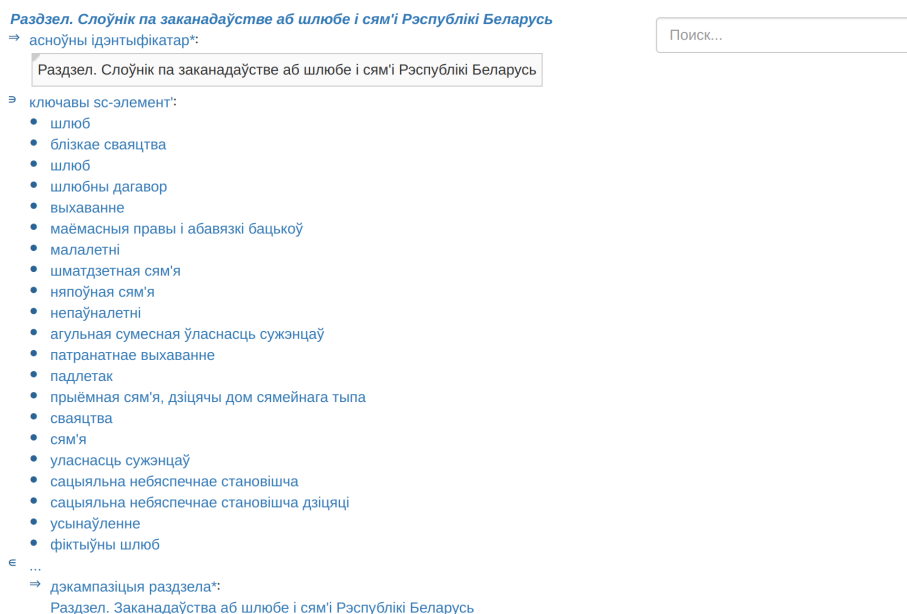


Рисунок 3.8 – Раздел. Словарь по законодательству о браке и семье Республики Беларусь

Каждое понятие содержит идентификаторы на русском, белорусском и английском языках, определение либо определения, а также ссылку на документ, в котором определено данное понятие (рис. 3.9).

подросток

⇒ основной идентификатор*:

подросток

€ ключевой sc-элемент*:

- Раздел. Словарь по законодательству о браке и семье Республики Беларусь
- Несовершеннолетний в возрасте от четырнадцати до восемнадцати лет считается подростком.

€ ответ на вопрос на естественном языке*:

Запрос на естественном языке: 3 понятия на букву П

€ подросток

⇒ Кодекс Республики Беларусь о браке и семье

€ белорусское юридическое понятие

Поиск...

Рисунок 3.9 – Понятие: подросток

Юридические документы содержат в себе сущности статей данного документа, основной идентификатор (рис. 3.10).

Кодекс Республики Беларусь о браке и семье

⇒ основной идентификатор*:

Кодекс Республики Беларусь о браке и семье

€ ключевой sc-элемент*:

Раздел. Законодательные акты о браке и семье Республики Беларусь

⇒ Статья 204. Порядок регистрации рождения

⇒ Статья 52. Установление происхождения детей, родившихся в результате применения вспомогательных репродуктивных технологий

⇒ Статья 122. Порядок усыновления ребенка

⇒ Статья 205. Заявление о регистрации рождения

⇒ Статья 123. Учет детей, подлежащих усыновлению, и лиц, желающих усыновить детей

⇒ Статья 206. Срок подачи заявления о регистрации рождения

⇒ Статья 54. Запись о родителях ребенка

⇒ Статья 124. Недопустимость незаконных действий по усыновлению детей

⇒ Статья 207. Запись акта о рождении

⇒ Статья 55. Запись о родителях, если материнство и (или) отцовство не установлены

⇒ Статья 125. Лица, имеющие право быть усыновителями

⇒ Статья 208. Регистрация рождения ребенка, родившегося после смерти отца либо объявления его в судебном порядке умершим

⇒ Статья 56. Установление материнства в судебном порядке

⇒ Статья 126. Разница в возрасте между усыновителем и усыновляемым ребенком

⇒ Статья 209. Регистрация рождения ребенка, родившегося после расторжения брака или признания брака недействительным

⇒ Статья 57. Установление отцовства в судебном порядке

⇒ Статья 127. Согласие родителей на усыновление

⇒ Статья 210. Порядок регистрации заключения брака

⇒ Статья 58. Оспаривание записи о родителях

⇒ Статья 128. Усыновление ребенка без согласия родителей

⇒ Статья 211. Заявление о регистрации заключения брака

⇒ Статья 59. Семья

⇒ Статья 129. Согласие опекуна, попечителя на усыновление детей-сирот, детей, оставшихся без попечения родителей

⇒ Статья 212. Ознакомление лиц, вступающих в брак, с условиями и порядком регистрации заключения брака, правами и обязанностями супругов

⇒ Статья 60. Близкое родство

Поиск...

Рисунок 3.10 – Кодекс Республики Беларусь о браке и семье

Статьи юридических документов содержат в себе их именование, текст статьи, а также ссылку на документ, частью которого является данная статья (рис. 3.11).

Статья 204. Порядок регистрации рождения

⇒ основной идентификатор*:

Статья 204. Порядок регистрации рождения

Поиск...

€ ключевой sc-элемент:

Регистрация рождения производится в органах, регистрирующих акты гражданского состояния, по месту рождения детей или по месту жительства родителей либо одного из них. По желанию родителей органы, регистрирующие акты гражданского состояния, обеспечивают торжественную обстановку регистрации рождения.

← Кодекс Республики Беларусь о браке и семье

Рисунок 3.11 – Статья 204. Порядок регистрации рождения

3.4 Вывод

В результате курсового проектирования были разработаны фрагменты базы знаний персонального ассистента по юриспруденции, а именно были формализованы предметные области, законодательные акты и понятия, относящиеся к сфере законодательства о браке и семье. Таким образом, база знаний была пополнена двумя предметными областями, двадцатью понятиями, а также в полном объеме в базу знаний был погружен Кодекс Республики Беларусь о браке и семье. Также в рамках курсового проектирования был разработан набор решателей задач, способных отвечать на вопросы о законодательстве Республики Беларусь, поставленные на естественном языке.

ЗАКЛЮЧЕНИЕ

В рамках данного курсового проекта были разработаны фрагменты базы знаний персонального ассистента по юриспруденции – словарь белорусского законодательства и база законодательных актов в сфере законодательства о браке и семье. Работа в рамках проекта была сфокусирована на поддержание актуальности знаний в базе, а также на предоставление максимального количества информации о белорусском законодательстве.

В процессе выполнения курсового проекта также было разработано семейство решателей задач, предоставляющих функционал ассистента по юриспруденции – а именно, возможность получения ответов на запросы на естественном языке, построенных в произвольной форме.

Результатами данной работы стала иерархия разделов и предметных областей базы знаний, в базу знаний были добавлено более сотни сущностей, включающих в себя юридические понятия, предметные области, отношения, законодательные акты Республики Беларусь, а также их составляющие, пять решателей задач, из которых двое решают задачи взаимодействия с Telegram-интерфейсом, трое решают задачи нахождения ответов на вопросы, поставленные пользователем.

Ожидаемый эффект от проделанной работы заключается в создании удобного инструмента, при помощи которого пользователь, не обладающий образованием в сфере юриспруденции сможет легко получать необходимую ему справочную информацию о законодательстве Республики Беларусь.

Дальнейшее развитие системы может включать в себя расширение базы знаний ассистента на другие сферы законодательства Республики Беларусь и других стран, реализации новых решателей задач, способных отвечать на большее множество вопросов, решать прикладные юридические задачи.

В целом, разработанная система представляет собой важный первый шаг к созданию полноценного интеллектуального персонального ассистента по юриспруденции, такой продукт может быть полезным не только простому пользователю, но и юристам, малым бизнесам, государственным служащим, ученым в области юриспруденции, результаты разработки могут быть использованы в исследованиях и при разработке иных систем.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

[1] Дробязко, С. Г. Отрасли права и отрасли законодательства в правовой системе Республики Беларусь и их совершенствование / С. Г. Дробязко // Право и экономика. — 2010. — № 5. — с. 11–22. — Режим доступа: https://law.bsu.by/pub/31/Drobzyazko_20.pdf. — Дата доступа: 07.10.2023.

[2] Теория государства и права в вопросах и ответах: учебное пособие. Издание второе, исправленное и доработанное / А. Ф. Колотов [и др.]. — ООО ИПК «Университет», 2013. — Режим доступа: <http://www.oimsla.edu.ru/sites/default/files/f50E.pdf>. — Дата доступа: 07.10.2023.

[3] Гаврилова, Т. А. Базы знаний интеллектуальных систем / Т. А. Гаврилова, В. Ф. Хорошевский. — Питер, 2000. — 384 с. — Режим доступа: http://www.immsp.kiev.ua/postgraduate/Biblioteka_trudy/BazyZnanIntelSistemGavrilova2000.pdf. — Дата доступа: 03.10.2023.

[4] Барышев, М. В. Модели представления знаний экспертных систем / М. В. Барышев, И. Ю. Гатчин, Ю. А. Гатчин // Научно-технический вестник информационных технологий, механики и оптики. — 2006. — № 29. — Режим доступа: <https://cyberleninka.ru/article/n/modeli-predstavleniya-znaniy-ekspertnyh-sistem>. — Дата доступа: 03.10.2023.

[5] Semantic Web – W3C [electronic resource]. — Mode of access: <https://www.w3.org/standards/semanticweb/>. — Date of access: 03.10.2023.

[6] Каешко, А. И. Принципы интеграции содержимого RDF-хранилищ в проект OSTIS / А. И. Каешко, Д. Г. Колб // Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS-2014) : материалы IV междунар. науч.-техн. конф. (Минск, 20-22 февраля 2014 года) / редкол.: В. В. Голенков (отв. ред.) [и другие]. — 2014. — с. 447 – 452. — Режим доступа: <https://libeldoc.bsuir.by/handle/123456789/26257>. — Дата доступа: 03.10.2023.

[7] Шункевич, Д. В. Модели и средства компонентного проектирования машин обработки знаний на основе семантических сетей / Д. В. Шункевич // Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS-2013) : материалы III Междунар. научн.-техн. конф. (Минск, 21-23 февраля 2013г.) / редкол. : В. В. Голенков (отв. ред.) [и др.]. — 2013. — с. 269 – 280. — Режим доступа: <https://libeldoc.bsuir.by/handle/123456789/4366>. — Дата доступа: 03.10.2023.

[8] Хту, Каунг Мьят. Анализ языка Веб онтологии (owl) и семантическая веб-технология / Каунг Мьят Хту // Auditorium. — 2017. — № 4. — Режим доступа: <https://cyberleninka.ru/article/n/analiz-yazyka-veb-ontologii-owl-i-semanticheskaya-veb-tehnologiya>. — Дата доступа: 03.10.2023.

[9] Кафтанников, И. Л. Перспективы использования web-онтологий в учебном процессе / И. Л. Кафтанников, С. Е. Коровин // ОТО. — 2003. — № 3. — Режим доступа: <https://cyberleninka.ru/article/n/perspektivy-ispolzovaniya-web-ontologiy-v-uchebnom-protssesse>. — Дата доступа: 03.10.2023.

[10] Голенков, В. В. Открытая технология онтологического проектирования, производства и эксплуатации семантически совместимых гибридных интеллектуальных компьютерных систем / В. В. Голенков, Н. А. Гулякина, Д. В. Шункевич. — Бестпринт, 2021. — 690 с. — Режим доступа: <https://libeldoc.bsuir.by/handle/123456789/45813>. — Дата доступа: 03.10.2023.

[11] ЭТАЛОН-ONLINE - Описание [электронный ресурс]. — Режим доступа: <https://etalonline.by/spravka/o-sisteme/>. — Дата доступа: 03.10.2023.

[12] Информационно-правовая система Нормативка.by [электронный ресурс]. — Режим доступа: <https://normativka.by/>. — Дата доступа: 07.10.2023.

[13] AI Lawyer - instant AI-powered legal help. — Mode of access: <https://ailawyer.pro/>. — Date of access: 10.11.2023.

[14] Docker Documentation – Overview [electronic resource]. — Mode of access: <https://docs.docker.com/get-started/overview/>. — Date of access: 07.10.2023.

[15] Java client library for SC machine. — Mode of access: <https://github.com/rastsislaux/jesc/>. — Date of access: 07.10.2023.

[16] Kotlin Documentation – Get started with Kotlin [electronic resource]. — Mode of access: <https://kotlinlang.org/docs/getting-started.html/> — Date of access: 07.10.2023.

[17] WitAI Docs [electronic resource]. — Mode of access: <https://wit.ai/docs>. — Date of access: 16.11.2023.