



Руководство пользователя

«Контент-менеджер»

Платформа CraftTalk cloud

Дата последней редакции 27.07.2021

2021

Содержание

1	Возможности системы	8
1.1	Как это работает	8
1.2	Сценарии использования	9
1.3	Проекты в платформе CraftTalk cloud	10
1.4	Компоненты платформы CraftTalk cloud	10
2	Рабочее место	12
2.1	Вход в систему	12
3	База знаний	13
3.1	Вкладка «Записи Базы знаний»	14
3.2	Добавление записи (обогащение Базы знаний)	18
3.3	Раздел «Общая информация»	19
3.4	Раздел «Примеры вопросов»	21
3.5	Раздел «Ответы»	25
3.6	Раздел «Слоты»	36
3.6.1	Пример создания статьи для распознавания товара в социальной сети Вконтакте	49
3.7	Раздел «Настройки» (создание статьи БЗ - настройки)	51
3.8	Пример создания нескольких ответов для нескольких каналов в рамках одной статьи	53
3.9	Активация новой записи в Базе знаний	55
3.10	Редактирование записи	58
3.11	Редактирование записи – «Общая информация»	59
3.12	Редактирование записи – «Ответы»	59
3.12.1	Редактирование записи – «Ответы». Синонимы для кнопки	62
3.13	Редактирование записи – «Размеченные сообщения»	65
3.14	Редактирование записи – «Примеры вопросов»	66
3.15	Редактирование записи – «Похожие вопросы»	67
3.16	Редактирование записи – «Слоты»	69
3.17	Активация отредактированной записи	69
3.18	Вкладка «Тематики»	70

3.19 Вкладка «Типы записей базы знаний»	72
3.20 Вкладка «Слоты»	73
3.20.1 Слот типа словарь. Функционал синонимов.	76
3.20.2 Аргументация синонимами	79
3.21 Вкладка «Стикеры»	81
3.21.1 Прием стикеров	82
3.21.2 Отправка стикеров	83
3.21.3 Добавление стикера для канала Telegram	83
3.21.4 Добавление стикера для канала ВКонтакте	85
3.21.5 Добавление стикера для канала Одноклассники	85
3.21.6 Добавление стикера для канала Facebook	85
3.22 Вкладка «Анкеты»	86
3.23 Описание классификаторов	87
3.23.1 Классификатор Customer_LOGREG	88
3.23.2 Классификатор Customer_GRU	89
3.23.3 Классификатор Customer_KNN	90
3.23.4 Классификатор customer_WMD	90
3.23.5 Рекомендации по переобучению классификаторов	92
3.23.6 Сбор разметки для классификаторов	92
3.24 Вкладка «Управление классификаторами»	93
3.25 Смена привязки к статье БЗ и Обучение Бота	95
4 Сценарии	97
4.1 Возможности Чат-бота	97
4.2 Распознавание сущностей и заполнение слотов	98
4.3 Ветвления диалога	99
4.4 Общее описание	100
Приложение 1. Классификаторы	107
1.1 Обучение чат-бота (классификатор WMD)	107
1.2 Примеры вопросов не на языке классификатора. Английские слова в примерах для обучения	110
1.3 Обучение классификаторов. Эмбединги. Переобучение эмбедингов	111

1.4 Описание классификатора KNN	112
1.5 Время и скорость обучения классификаторов	116
1.6 Список стоп-слов чат-бота	116
Приложение 2. Добавление ответов для различных каналов	118
2.1 Добавление ответов для различных каналов в рамках одной статьи	118
2.2 Создание слота “Каналы”	118
2.3 Создание нескольких ответов для нескольких каналов в рамках одной статьи	119
Приложение 3. Пример создания сценариев	121
Приложение 4. Реализация интеллектуального бота	126
Пример обучения чат-бота с классификатором WMD	126
Приложение 5. Ограничения Markdown по каналам Платформы	130
Лист регистрации изменений	134

Определения и сокращения

API – программный интерфейс приложения, интерфейс прикладного программирования. Описание способов (набор классов, процедур, функций, структур или констант), которыми одна компьютерная программа может взаимодействовать с другой программой. Используется программистами при написании всевозможных приложений.

Агент – подпрограмма, выполняющая обработку данных. Программами агентами являются: начальное приветствие, Бот, Агент маршрутизации по Операторам, рабочее место Оператора и т.п.

Аккаунт – это совокупность данных, которая описывает пользователя.

База знаний – база данных, содержащая правила вывода информации, сведения о продуктах и услугах, а также результаты обработки предыдущих сообщений.

ИИ Бот – специальная программа, используемая в мессенджере и веб-чате, являющаяся частью электронного виртуального ассистента (ЭВА), выполняющая поиск и извлечение из Базы знаний вариантов ответов на вопрос Обратившегося для автоматического направления ответа непосредственно Обратившемуся через виджет или мессенджер.

Виджет – класс вспомогательных мини-программ — графических модулей, которые размещаются в рабочем пространстве соответствующей родительской программы и служат для решения отдельных рабочих задач: авторизации клиента, обмена сообщениями, получение информации из сети интернет.

Виджетом называют не само приложение, а графический или текстовый элемент, который открывает к нему доступ или запускает работу приложения.

Группа операторов – несколько Операторов, объединенных по квалификационному признаку, т.е. по способности обрабатывать определенные типы обращений.

Идентификация – совокупность мероприятий по установлению сведений о клиентах, их представителях.

Клиент – физическое лицо, пользующееся услугами Контакт-центра компании.

Контакт-центр (КЦ) – подразделение компании, осуществляющее обработку обращений потенциальных и действующих клиентов.

Мессенджер – информационно-коммуникационный сервис обмена мгновенными сообщениями.

Обратившийся – физическое лицо, в том числе представитель юридического лица, направляющее сообщение при помощи дистанционного канала коммуникации.

Обращение – набор сообщений, не носящих рекламный характер и отправленных Обратившимся для получения консультации по услугам и продуктам компании, передачи жалоб, отзывов и предложений.

Оператор – работник Контакт-центра, осуществляющий обработку входящих обращений, поступающих в Контакт-центр по единым телефонным номерам и другим дистанционным каналам (электронная почта, веб-чат, мессенджер, формы обратной связи, звонок с сайта, обратный звонок, размещенные на сайте компании в сети интернет).

ПО – программное обеспечение.

Пользователь – специалист компании, принимающий участие в эксплуатации программного обеспечения.

Сессия – сеанс обмена сообщениями между Обратившимся и Оператором Контакт-центра. Сессия начинается с момента маршрутизации сообщения Оператору из общей очереди и продолжается до её завершения Обратившимся, оператором Контакт-центра или по тайм-ауту, определяемому настройками ПО.

Супервизор – работник Контакт-центра, осуществляющий контроль за деятельностью Операторов и групп операторов, в том числе с целью оперативного управления.

Электронный виртуальный ассистент (ЭВА) – информационная система, предназначенная для извлечения из Базы знаний вариантов ответов на вопрос Обратившегося и вывода их на диалоговое окно в удобном формате в помощь Оператору при подготовке им ответа непосредственно Обратившемуся либо автоматически при помощи Бота. При подготовке варианта ответа или подсказки ЭВА использует семантический, лингвистический и морфологический анализ текста сообщения.

1 Возможности системы

Платформа CraftTalk cloud - это универсальное решение для обслуживания клиентов, предназначенное для контакт-центров и внутренних нужд компании, таких как служба поддержки персонала и IT. Платформа использует совместную работу людей и искусственного интеллекта для создания высококачественного клиентского сервиса в чате и голосовых каналах.

С платформой CraftTalk cloud Вы можете начать предоставлять клиентское обслуживание всего за 1 день с использованием агентов-людей и базовой автоматизации. Уровень автоматизации будет расти со временем, поскольку искусственный интеллект (ИИ) будет учиться у людей. Автоматизация также быстро адаптируется к изменениям в структуре трафика, мгновенно обучаясь обработке новых, ранее неизвестных запросов. В наших реальных случаях уровень автоматизации достигает от 50 до 70% через 3 месяца для контакт-центров, которые обслуживают до 7 тысяч различных тем. Время адаптации к новым «горячим» темам может составлять от одного до нескольких часов.

Платформа CraftTalk cloud достигает этих результатов, используя самые передовые технологии искусственного интеллекта для обеспечения высокой эффективности операций, обеспечивая мгновенное обучение искусственного интеллекта на основе взаимодействия людей и не требуя дорогостоящих специалистов по IT / машинному обучению для настройки и обслуживания служб искусственного интеллекта.

1.1 Как это работает

Процесс взаимодействия с пользователями в платформе CraftTalk cloud выглядит следующим образом:

- пользователь общается с компанией через текстовый или голосовой канал;
- ИИ пытается обработать запрос простым ответом или сложным взаимодействием, которое также может включать интеграцию;
- если ИИ не может найти информацию он может перенаправить запрос живым агентам (операторам);
- операторам в процессе обработки обращений помогает ИИ, и они могут использовать навыки чат-бота (интеллектуального ассистента) для выполнения сложных задач.

С платформой CraftTalk cloud Вам не нужно глубоко погружаться в концепции машинного обучения или терминологию понимания естественного языка, собирать наборы данных для обучения, подключать несколько платформ для ИИ и операторов и т.д. Вместо этого Вы заполняете статьи в уже знакомой базе знаний, сосредотачиваясь на том, как форматировать красивые сообщения для мессенджеров. Выберите, какая информация Вам нужна от пользователя, а под капотом платформа CraftTalk cloud будет использовать эти данные для создания чат-бота на базе искусственного интеллекта, который будет учиться у операторов, помогать людям выполнять сложные бизнес-задачи.

Добавить новый навык обслуживания клиентов для бота и человека в платформе CraftTalk cloud так же просто, как добавить новую статью в базе знаний.

1.2 Сценарии использования

Типичные задачи, которые можно решить с помощью платформы CraftTalk cloud:

- организовать обслуживание клиентов в текстовых каналах, таких как чат на сайте, в мобильных приложениях, мессенджерах, социальных сетях;
- организовать обслуживание в голосовых каналах, таких как голосовые помощники и телефон;
- отвечать на запросы клиентов в текстовых и голосовых каналах с помощью ИИ и операторов;
- организовать самообслуживание в чатах и голосовых помощников с такими функциями, как получение баланса, осуществление платежей, изменение тарифов;
- создать дружелюбную IT-службу поддержки и HR-помощников;
- создать внутрикорпоративную базу знаний / вики для сотрудников с интеллектуальным самообучающимся поиском на основе ИИ и богатым контентом;
- создать базу знаний для операторов контакт-центра с интеллектуальным поиском на основе ИИ, который самообучается на основе взаимодействия с операторами;
- создать интеллектуальных ассистентов (ИИ-ботов) для клиентов и сотрудников;
- создать навыки для интеллектуальных ассистентов.

Ключевая концепция: платформа CraftTalk cloud - это все, что Вам нужно для создания полностью работающего многоканального контакт-центра на базе ИИ с нуля или для перехода с существующей платформы с минимальными усилиями для повышения качества обслуживания и снижения стоимости запроса.

1.3 Проекты в платформе CraftTalk cloud

Для начала работы с платформой CraftTalk cloud Вы создаете новый проект и приглашаете в него пользователей по электронной почте. Все данные проекта не зависят от других проектов. Пользователь может быть добавлен в несколько проектов с разными ролями и при необходимости переключаться между проектами.

Важная особенность платформы CraftTalk cloud заключается в том, что оператора можно добавить в очереди в нескольких проектах. Таким образом они могут получать задачи из нескольких проектов. Это не требует переключения проектов со стороны операторов. Все настройки, такие как приоритет очереди и размер задачи, работают так же, как если бы очереди были из одного проекта.

Если у Вас есть выделенный экземпляр платформы CraftTalk cloud, Вы можете полностью контролировать всех пользователей, включая добавление или блокировку входа пользователей. В противном случае Вы можете контролировать доступ пользователей к вашему проекту и их роли в Вашем проекте, но не можете контролировать активность пользователя за пределами границы Вашего проекта.

1.4 Компоненты платформы CraftTalk cloud

Платформа CraftTalk cloud предоставляет следующие основные компоненты, которые наиболее эффективны при совместной работе, но также могут использоваться по отдельности:

- **Модуль базы знаний** - решение нового поколения для организации знаний компании, данных для обучения искусственного интеллекта, бизнес-процессов и интеграции.

- **Модуль интеллектуальный чат-бот** - технология искусственного интеллекта, основанная на новейших алгоритмах NLU, базирующихся на глубоком обучении, для автоматизации задач обслуживания клиентов. Вы можете создавать неограниченное количество интеллектуальных чат-ботов.

- **Модуль рабочее место оператора** - решение для контакт-центра, которое включает в себя интеллектуальное управление очередями и рабочее место для операторов с поддержкой интеллектуального ассистента.

- **Модуль коммуникационная система** - омниканальная шина обмена сообщениями, которая объединяет текстовые и голосовые каналы (например, чат на веб-сайте, в мобильных приложениях, мессенджерах, социальных сетях, голосовых помощниках и голосовых вызовах) и использует операторов и интеллектуальных чат-ботов для обработки запросов. Она также хранит всю коммуникацию и информацию о клиентах.

- **Модуль статистика** - инструмент для создания гибких панелей мониторинга для отслеживания и настройки каждого аспекта клиентского сервиса.

2 Рабочее место

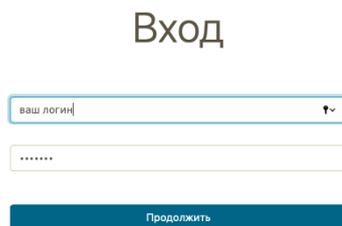
Автоматизированное рабочее место (далее АРМ) предоставляет необходимые функции для всех пользователей платформы – от Оператора до Супервайзера. Возможность доступа к различным функциям регулируется наличием необходимых ролей у пользователя и настройками пользователя по маршрутизации. Доступ к рабочему месту предоставляется только после авторизации.

Взаимодействие пользователей с автоматизированным рабочим местом осуществляется при помощи пользовательского интерфейса Системы «CraftTalk», реализованного в виде веб-клиента, вызываемого по ссылке. Рекомендуется использовать браузеры: Chrome (версии не ниже 71.0.3578.98) и IE (не ниже 11 версии).

2.1 Вход в систему

Для входа на рабочее место необходимо (рис. 2.1):

- в браузере перейти по адресу, предоставленному системным администратором;
- использовать зарегистрированные логин и пароль.



Вход

ваш логин

.....

Продолжить

Рисунок 2.1

После удачной авторизации в системе открывается основное окно интерфейса «CraftTalk» на вкладке «Работа с обращениями».

Функционал работы со статистикой и поиск обращений подробно описан в Руководстве пользователя «Супервизор»

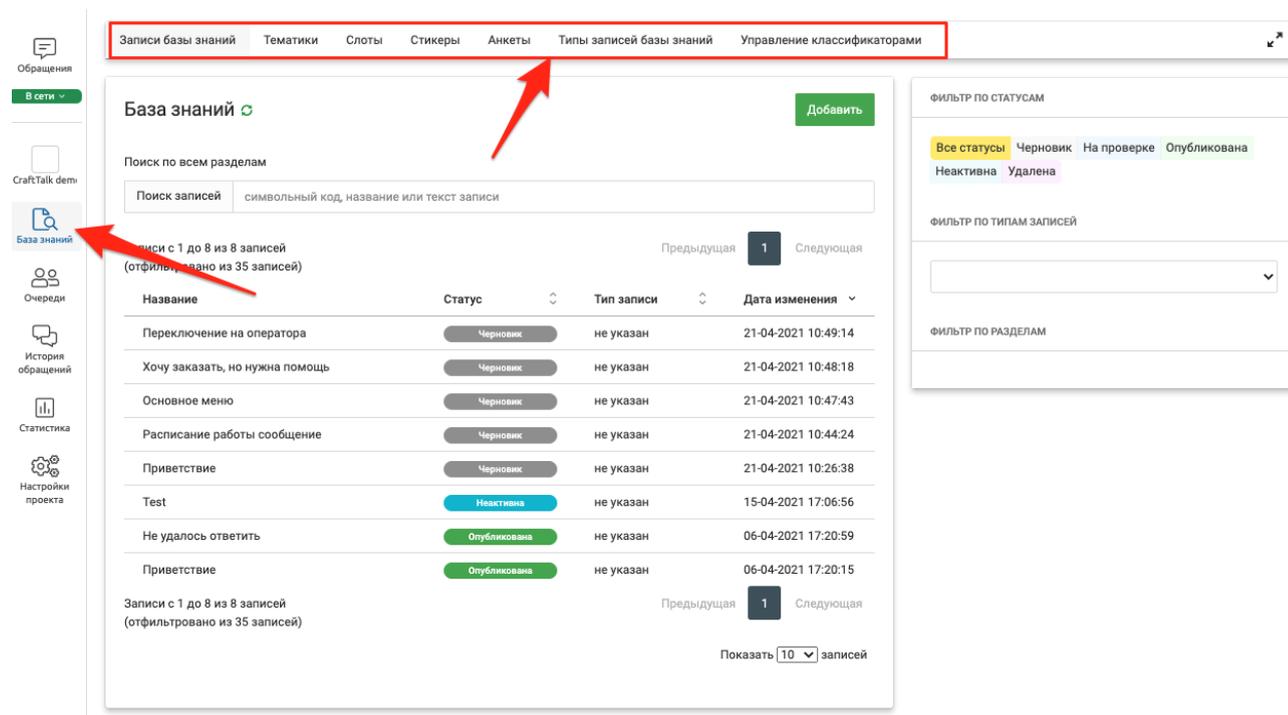
3 База знаний

В разделе «База знаний» создаются и редактируются навыки – новые статьи Базы знаний.

В пункте меню «База знаний» доступен редактор, который позволяет искать, а также заносить новые и редактировать существующие статьи Базы знаний (рис. 3.1).

Вверху главного меню «База знаний» расположены вкладки:

- «Записи базы знаний»;
- «Тематики»;
- «Стикеры»;
- «Слоты»;
- «Анкеты»;
- «Типы тематик»;
- «Типы записей базы знаний»;
- «Управление классификаторами».



База знаний

Поиск по всем разделам

Поиск записей (символьный код, название или текст записи)

Записи с 1 до 8 из 8 записей (отфильтровано из 35 записей)

Название	Статус	Тип записи	Дата изменения
Переключение на оператора	Черновик	не указан	21-04-2021 10:49:14
Хочу заказать, но нужна помощь	Черновик	не указан	21-04-2021 10:48:18
Основное меню	Черновик	не указан	21-04-2021 10:47:43
Расписание работы сообщение	Черновик	не указан	21-04-2021 10:44:24
Приветствие	Черновик	не указан	21-04-2021 10:26:38
Test	Неактивна	не указан	15-04-2021 17:06:56
Не удалось ответить	Опубликована	не указан	06-04-2021 17:20:59
Приветствие	Опубликована	не указан	06-04-2021 17:20:15

Записи с 1 до 8 из 8 записей (отфильтровано из 35 записей)

Показать 10 записей

ФИЛЬТР ПО СТАТУСАМ

Все статусы Черновик На проверке Опубликована Неактивна Удалена

ФИЛЬТР ПО ТИПАМ ЗАПИСЕЙ

ФИЛЬТР ПО РАЗДЕЛАМ

Рисунок 3.1

Информация из Базы знаний доступна пользователю при работе с обращениями.

Записи базы знаний используются:

- в качестве справочников для Оператора (в том числе могут содержать вложения в виде файлов и изображений);
- в качестве готовых ответов на типичные вопросы Обратившихся – могут использоваться Оператором, а также Ботом для автоматического ответа;
- как готовые ответы, для ответа на которые Бот может запрашивать дополнительную информацию и отвечать, исходя из ответов Обратившихся, либо передавать эту информацию сторонним агентам (интеграциям);
- для создания шаблонов исходящих обращений.

Визуальная рабочая область поделена на две части (рис. 3.2). Справа находятся фильтры для поиска записей в Базе знаний и классификаторы, которые позволяют уточнить поиск записей по Базе знаний. В центре экрана – строка поиска записей и таблица с перечнем записей, занесённых в Базу знаний и кнопка обновления информации .

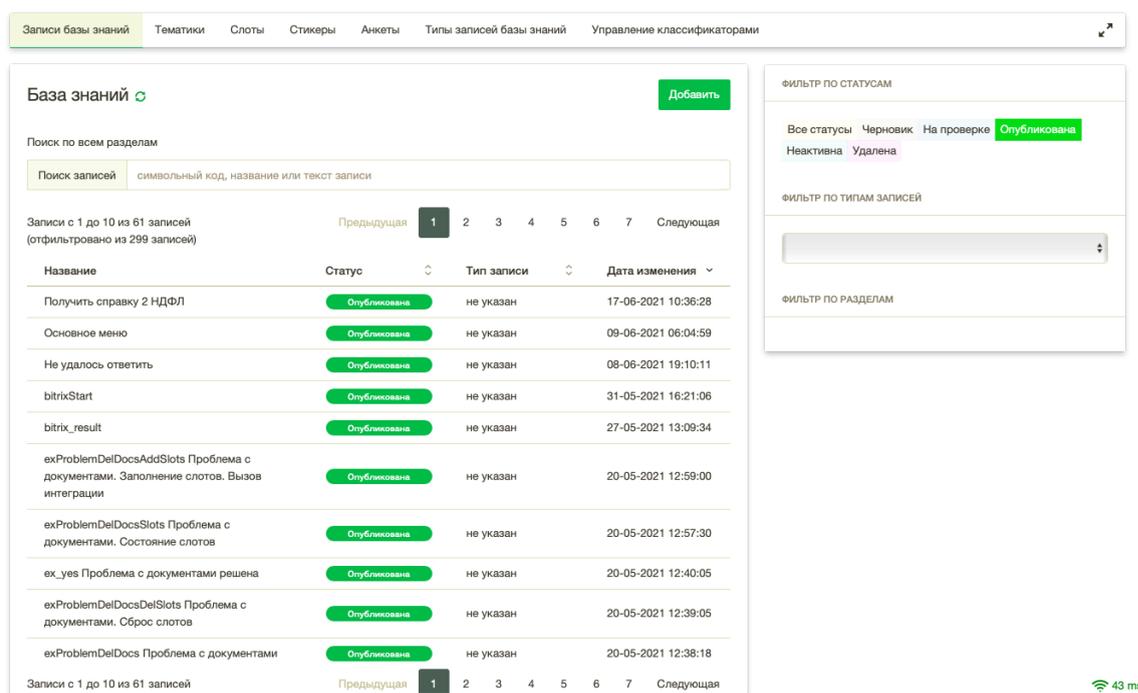


Рисунок 3.2

3.1 Вкладка «Записи Базы знаний»

Для поиска в Базе знаний пользователи могут использовать полнотекстовые запросы, при этом при наборе запроса система выводит отсортированные по введённому тексту записи (рис. 3.3).

Записи базы знаний Тематики Слоты Стикеры Анкеты Типы записей базы знаний Управление классификаторами

База знаний  Добавить

Поиск по всем разделам

Поиск записей

Записи с 1 до 2 из 2 записей (отфильтровано из 299 записей) Предыдущая 1 Следующая

Название	Статус	Тип записи	Дата изменения
Получить справку 2 НДФЛ	Опубликована	не указан	17-06-2021 10:36:28
Налоговый вычет. Текст ответа	Опубликована	Клиника	03-05-2021 14:28:45

Записи с 1 до 2 из 2 записей (отфильтровано из 299 записей) Предыдущая 1 Следующая

Показать записей

Рисунок 3.3

Для удобства пользования системой по умолчанию записи в таблице разбиваются на страницы и выводятся по 10 строк. Под таблицей есть выпадающий список, позволяющий настраивать количество выводимых на странице записей. Рядом расположен механизм перехода по страницам таблицы (кнопки “Предыдущая”, “Следующая”). Номер текущей страницы выделен фоном.

Записи в таблице можно отсортировать по любому видимому столбцу, кроме «Названия»: «Статус», «Тип записи», «Дата изменения».

Столбец, по которому в текущий момент времени отсортирована таблица, отмечен направленной вверх или вниз галочкой. Направление галочки зависит от способа сортировки.

Для нахождения интересующей записи Базы знаний достаточно ввести в строку поиска название статьи Базы знаний, ее код, или текст записи, а также ключевые слова, синонимы, которые предварительно заносятся во вкладку «Примеры вопросов» при создании новой статьи Базы знаний. По мере ввода системой будут отфильтровываться не удовлетворяющие введенному условию записи.

Информация в Базе знаний может быть отфильтрована по типам записей, статусу, тематикам и дате изменений.

Справа расположена таблица фильтров для сортировки записей по статусу, по типам записи, разделам, тематикам (рис. 3.4). Для сортировки поиска, например по статусам можно отдельно отобразить «черновик», «опубликована» (статьи, которые доступны всем сотрудникам Контакт центра), «неактивна» (эти статьи не видны сотрудникам Контакт центра, но их в любой момент можно активировать т.е. опубликовать), «на проверке» (статьи, которые в данный момент проверяются администратором перед публикацией в общий доступ) или «Удалена». Для того, чтобы осуществить поиск по всем статусам, можно выбрать соответствующую категорию «Все статусы». Возврат в просмотр не отфильтрованных данных осуществляется нажатием «Очистить фильтр».

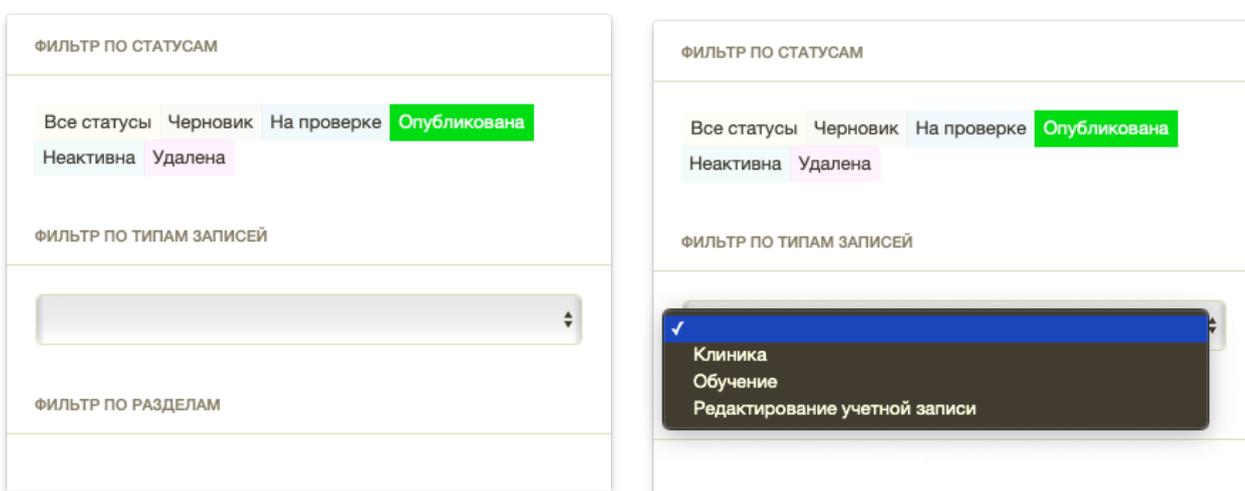


Рисунок 3.4

Во вкладке есть возможности - добавить запись и редактировать запись. Механизм добавления и редактирования практически одинаков. Во всех случаях заполняются поля открытой формы или вносятся изменения.

Нажатие на строке таблицы с записью открывает форму редактирования и удаления статьи базы знаний.

Типичная статья БЗ, используемая в качестве справочника для оператора, состоит из следующих элементов:

- название;
- символьный код статьи;
- раздел (где сохраняется тематика, к которой относится данная статья Базы знаний);

- тип записи (используется для разграничения записей, типичный вариант: «ФЛ» - физические лица и «ЮЛ» – юридические лица);

- статус.

К каждой статье привязаны:

- ответы;

- размеченные сообщения;

- примеры вопросов;

- похожие вопросы;

- слоты;

- настройки.

Каждый ответ может содержать помимо текста:

- вложения в виде файлов или изображений;

- «кнопки», которые пользователь может использовать для быстрого перехода к другой сопутствующей информации (к другим статьям БЗ).

Ответ может зависеть от значения слота, если он задействован в ответе.

Помимо явно указанных формулировок, интеллектуальный помощник и ИИ бот используют для обучения информацию, накопленную в процессе работы операторов с БЗ (выбора оператором статей для отправки в качестве ответов на вопрос) и в процессе ручной разметки диалогов по истории обращений.

Нажав на любую запись, можно просмотреть следующую информацию записи (рис. 3.5):

- Общая информация.

- Ответы (содержит варианты ответов).

- Размеченные сообщения.

- Примеры вопросов (Ожидаемые вопросы пользователей используются интеллектуальным помощником Оператора для подсказок статей из БЗ, а также ИИ ботом для автоматических ответов).

- Похожие вопросы.

- Слоты (разделение базы знаний на типы вопросов).

- Настройка (позволяет управлять выдачей сообщений).

Редактирование записи

id: 8aff0fc1-0153-48fb-9a05-5582d45d00b5

Общая информация | Ответы | Размеченные сообщения | Примеры вопросов | Похожие вопросы | Слоты | Настройки

Название

Символьный код

Разделы

Тип записи

Статус: Черновик

21.04.2021 10:49:14

● Черновик

abarkulova@crafttalk.ru

06.04.2021 17:20:59

● Опубликована

abarkulova@crafttalk.ru

Сравнить

06.04.2021 17:20:52

06.04.2021 17:20:59

● Черновик

abarkulova@crafttalk.ru

Сравнить

06.04.2021 17:19:26

06.04.2021 17:20:59

● Опубликована

system

Сравнить

06.04.2021 17:19:25

06.04.2021 17:19:26

● На проверке

system

Сравнить

06.04.2021 17:19:24

06.04.2021 17:19:25

● Черновик

system

Сравнить

Удалить навсегда
Отмена
Отправить на проверку
Опубликовать
Сохранить черновик

Рисунок 3.5

3.2 Добавление записи (обогащение Базы знаний)

Для создания новой статьи Базы знаний требуется нажать кнопку «Добавить», расположенную справа от строки поиска записи.

Создание записи

Общая информация | Ответы | Примеры вопросов | Слоты | Настройки | Граф интенгов

Название

Символьный код

Разделы

Тип записи

Статус:

Отмена
Создать черновик

Рисунок 3.6

Откроется вкладка «Создание записи» (рис. 3.6), где доступно создание новой записи и содержатся следующие разделы:

- Общая информация;

- Ответы;
- Примеры вопросов;
- Слоты;
- Настройка.

3.3 Раздел «Общая информация»

В этом разделе заполняются (см. рис. 3.6):

- Название;
- Символьный код;
- Разделы;
- Тип записи;
- Статус.

Обязательным к заполнению полем является поле “Символьный код”. Система допускает к вводу в поле символьного кода цифры, латинские и русские буквы, другие символы (например, нижнее подчеркивание, тире) и их комбинации. В случае пустого поля “Символьный код” при сохранении созданной записи будет выдано сообщение в левом нижнем углу экрана о невозможности создания новой записи в Базе знаний по причине отсутствия символьного кода. Символьный код после создания статьи нельзя отредактировать.

Поле «Название» желательно также не оставлять пустым, чтобы можно было идентифицировать запись в таблице просмотра.

Поле «Разделы» проставляется для разделения статей БЗ, здесь сохраняется тематика, к которой относится данная статья БЗ.

При наборе текста предполагаемого названия в строке появляется таблица вариантов тематик и подтематик.

Ошибочно введенную тематику можно удалить, нажав на красную кнопку с белым крестиком.

Поле «Тип записи» используется для разграничения статей (рис. 3.7), например:

- ФЛ – физические лица;
- ЮЛ – Юридические лица;
- Прочие – прочие навыки (типа Приветствие, Прощания и общие вопросы).



Рисунок 3.7

Статус проставляется автоматически и показывает статус, в котором в данный момент находится запись.

При редактировании уже существующий статьи БЗ в нижней части редактора во всех вкладках присутствуют кнопки (рис. 3.8), позволяющие осуществлять соответствующие действия при редактировании.

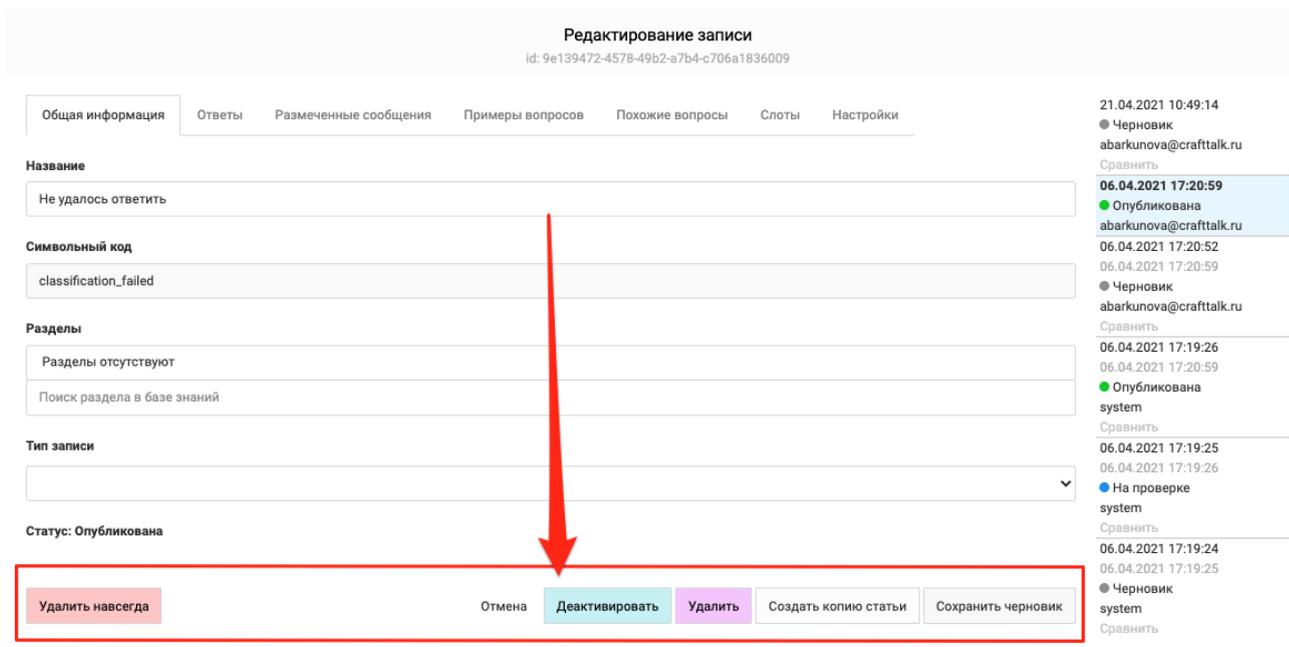


Рисунок 3.8

Это кнопки для публикации (активации) записи или её удаления, кнопка «Удалить» и «Деактивировать» без удаления, создания копии статьи для внесения изменений и подготовки новой версии. Публикация (активация) новой и отредактированной записи описана в параграфе «Активация записи».

Далее заполняются следующие разделы.

3.4 Раздел «Примеры вопросов»

В это поле вводятся различные варианты вопросов и фраз клиента, которые он может задать в диалогах на данную тему. Здесь добавляются ключевые слова и синонимы, по которым осуществляется поиск статей в Базе знаний. То есть в индексации участвуют не только названия статей, но и сохраненные примеры вопросов. Для добавления вопросов необходимо заполнить строку «Список вопросов» (рис. 3.9).

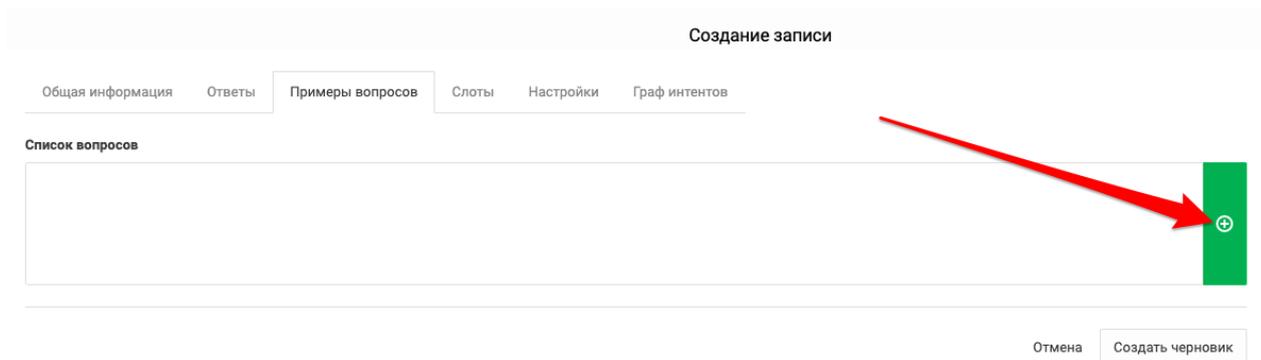


Рисунок 3.9

Для добавления текста вопросов необходимо нажать на кнопку «Примеры вопросов». Далее заполняем «Список вопросов», разделяя каждый вопрос символом перевода строки (нажатием Enter). Количество вопросов неограниченно и может состоять из ожидаемых вопросов пользователей, различных формулировок и похожих вопросов.

Примеры вопросов пользователей используются ЭВА Оператора для подсказок статей из БЗ, а также Ботом для автоматических ответов.

Помимо явно указанных формулировок, ЭВА и Бот используют для обучения информацию, накопленную в процессе работы Операторов с БЗ (выбора Оператором статей для отправки в качестве ответов на вопрос) и в процессе ручной разметки диалогов по истории обращений.

Можно отредактировать приведённые примеры вопросов, отредактировав их текст в «Списке вопросов», где можно внести сразу несколько вопросов в «Список вопросов» (рис. 3.10).

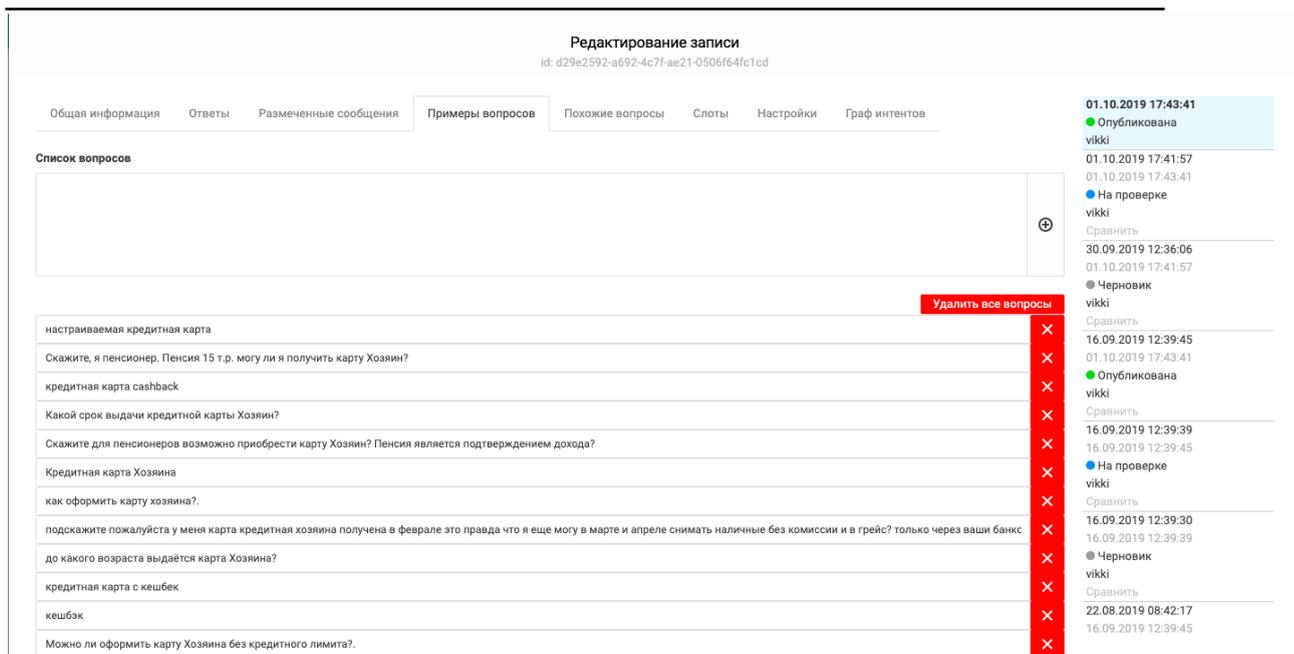


Рисунок 3.10

Для редактирования текста вопроса достаточно мышкой выделить вопрос и внести необходимые корректировки (рис. 3.11).



Рисунок 3.11

Добавление вопросов завершается нажатием кнопки  (см. рис. 3.9). Для удаления вопроса – нажать на кнопку  справа от вопроса (см. рис. 3.10).

Для лучшего обучения требуется введение как можно большего числа вопросов в различных формулировках, при этом не требуется введения вопросов в различных падежах и числе. Количество вопросов не ограничено.

Записи также можно деактивировать, удалить и создать копию (см. рис. 3.8).

Для разметки (примеров вопросов) необходимо вносить те фразы клиента, на которые будет отвечать статья.

Рекомендуется следующий процесс разметки данных для классификаторов:

1. Создание новой статьи.
2. Добавление характерных примеров (хотя бы 10 примеров). Например, для статьи «Зарплатная карта» требуются примеры вопросов про зарплатную карту,

и при первоначальной разметке акцент в фразах должен быть на словах зарплатная, зарплата, заработная плата ("мне нужна дебетовая карта для перечисления зарплаты", "у вас есть зарплатные карты", "нужна карта для з/пл", "для заработной карты какую карту посоветуете"). Это позволит классификатору WMD обучиться данной статье и начать отвечать на подобные формулировки.

3. Если от клиентов ранее в чате Оператору уже поступали вопросы по тематике новой статьи - рекомендуется подобрать подходящую разметку с помощью инструмента "Похожие вопросы". Данный инструмент анализирует неразмеченную историю диалогов и автоматически подбирает вопросы, похожие на уже указанные для статьи примеры. Необходимо отобрать подходящие вопросы с помощью "галочек" и нажать "сохранить отмеченные".

4. Дополнительно к автоматическому подбору инструмент "Похожие вопросы" позволяет подбирать подходящую разметку в истории по ключевым словам.

5. В дальнейшем в первую очередь разметка должна поступать от Операторов, при поступлении вопроса на них. Операторы должны преимущественно пользоваться инструментом "Самолетик" для ответа на вопрос, чтобы автоматически пополнять разметку статей.

6. Важно контролировать накапливаемую разметку с помощью вкладки "Размеченные сообщения" в статье при появлении сомнений в качестве ответов Бота на вопросы, связанные с определенной статьей.

7. В целом, чем больше разнообразных вопросов клиентов оказывается размечено, тем более качественно Бот сможет отвечать на вопросы клиентов и тем больше вопросов он будет закрывать самостоятельно.

Рекомендация!

- Не рекомендуется создавать разные статьи с очень близкой разметкой. Это затрудняет работу Бота, увеличивает сложность работы ответственного администратора со временем и, как правило, приводит к появлению неоднозначной разметки в процессе работы Операторов (так как они зачастую тоже не корректно используют такие статьи). Рекомендуется объединять такие статьи в одну статью.

В системе у агента AIAssist2 есть возможность уточнять запрос при необходимости. Это позволяет боту выводить возможные интенты, score которых выше порога BotResultClarificationThreshold в конфигурации классификатора, если не нашлось интенета, score которого превысил порог BotClassificationResultThreshold. Отображение осуществляется в виде кнопок, текст на которых - название интенета, в порядке уменьшения score. Количество интенетов, используемых при уточнении, задается настройкой "Максимальное количество интенетов для уточнения", но не менее 2 вариантов.

Есть возможность уточнять запрос даже при наличии ответа. Это позволяет боту уточнять интенет даже в том случае, если есть был получен классифицированный интенет. Работает только в совокупности с параметром "Уточнять запрос при необходимости".

Сообщение при уточнении намерения пользователя – это текст сообщения, которое бот будет отправлять пользователю при уточнении интенета.

Реализована возможность использовать двухшаговое уточнение. Это позволяет боту уточнять запрос пользователя при первой неудачной классификации. Счетчик сбрасывается после удачной классификации.

Идентификатор интенета для редиректа при уточнении в режиме "двухшаговое уточнение" – это идентификатор интенета, который будет использован ботом при использовании двухшагового уточнения.

- При объединении близких статей в одну статью рекомендуется формировать нейтральный ответ, подходящий под различные возможные вопросы по тематике статьи. Можно дополнительно добавлять связанные статьи без разметки и ссылаться на них с помощью механизма кнопок - это позволяет формировать краткие ответы на конкретные вопросы клиентов, при этом разметка формируется для основной статьи. Пример плохого ответа: "Нет, вы можете ..." - даже если статья отвечает на вопрос клиента, совсем не обязательно вопрос клиента подразумевает ответ с отрицанием.

- Крайне важно контролировать отсутствие одинаковой разметки у разных статей, используемых одним Ботом!

- При добавлении новой фразы в качестве примера, всегда нужно думать: "Действительно ли эта статья, куда я добавляю пример, ответит на вопрос неизвестного клиента, если он первым сообщением напишет данную фразу?".

3.5 Раздел «Ответы»

«Тип текста ответа» устанавливается (рис. 3.12):

- «HTML» - можно использовать, если создается статья для веб-чата (все другие каналы не поддерживают в полной мере html, поэтому там могут появиться ошибки) и для информации Операторов.

Для ответов в канал веб-чат можно использовать или тип ответа html, или тип ответ простой текст, но следует учитывать, что в случае использования форматирования в веб-чате, оно останется только при использовании типа ответа html, для всех остальных каналов можно использовать тип ответа простой текст или markdown, а для ответов, содержащих команды - только тип ответа "простой текст".

- "Простой текст" - для всех каналов (в том числе и без указания канала через слот "Каналы"). Могут использоваться емоji.

- «Markdown» - для обеспечения унифицированного форматирования сообщений в мессенджерах и социальных сетей. Основной markdown синтаксис, расширенный дополнительным блоком ответа и блоком кнопок.

При редактировании статьи БЗ могут быть изменены, добавлены или удалены ответы, удалены и добавлены кнопки, изменены значения слотов.

Для внесения текста ответов необходимо перейти в поле «Ответ» и далее ввести текст ответа в редакторе текста. Следует помнить, что количество ответов к одной статье Базы знаний неограниченно.

Общая информация | **Ответы** | Примеры вопросов | Слоты | Настройки | Граф интенгов

Ответы

Ответ ✕

Тип текста ответа

Простой текст HTML Markdown

Добавить шаблон ▼

Рисунок 3.12

Ответ вводится в специальное поле «Ответы» (рис. 3.13).

Укорачивание ссылок:

Введите адрес ссылки Укоротить Скопировать

Вложения:

Введите адрес документа Прикрепить

Кнопки:

Поиск записи в базе знаний ⊕ Добавить кнопку

Значения слотов:

▼ ⊕ Добавить слот

⊕ Добавить ответ

Отмена Создать черновик

Рисунок 3.13

Нажатие этой кнопки приводит к появлению плашки со ссылкой «Ответ», которая разворачивает форму ввода нового ответа (рис. 3.14) (состоит из двух страниц – ввода нового ответа и редактирования существующего), нажать «Добавить ответ».

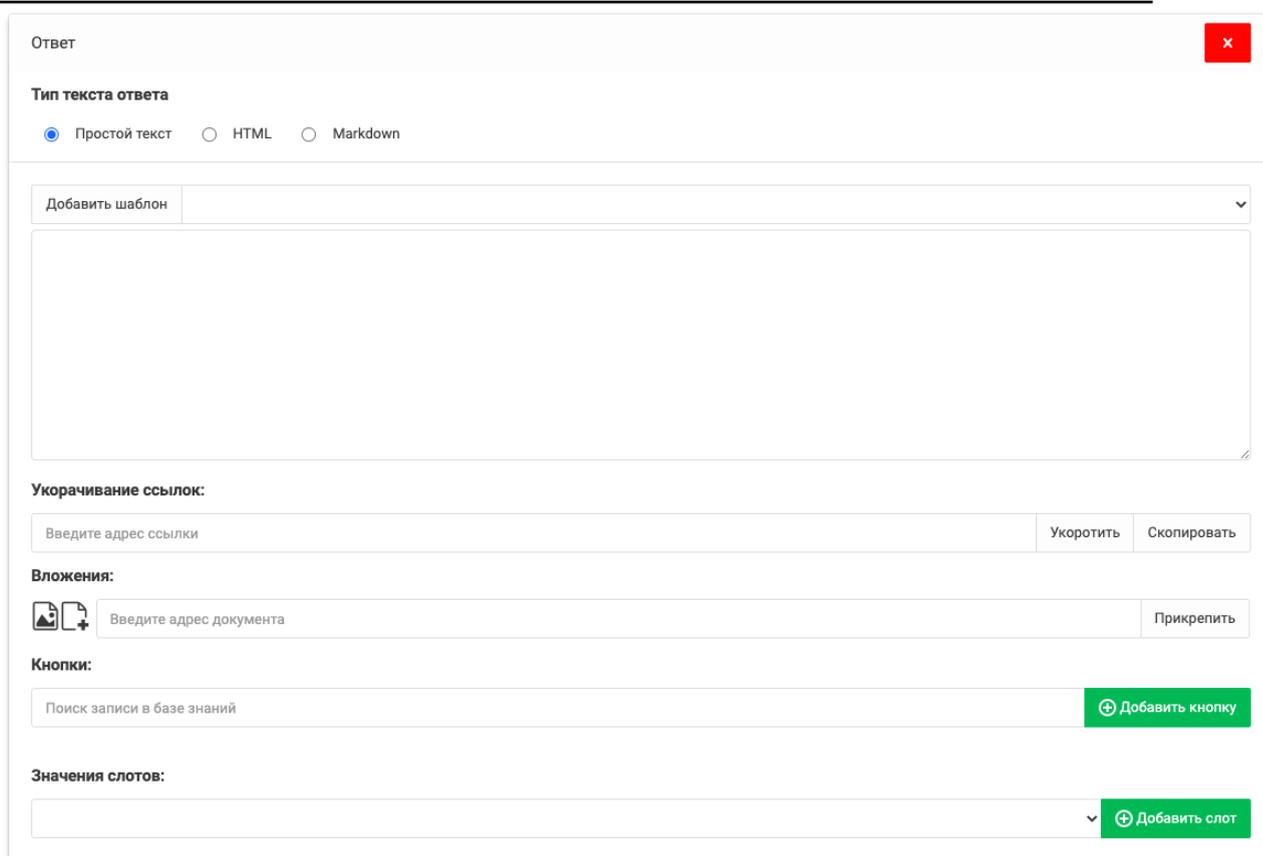


Рисунок 4.14

В случае необходимости ответ можно удалить из записи. Для удаления ответа надо нажать на кнопку удаления  справа от ответа и подтвердить удаление ответа в открывшемся диалоговом окне.

Каждый ответ может содержать помимо текста (см. рис. 3.14):

- вложения в виде файлов или изображений;
- «кнопки», которые пользователь может использовать для быстрого перехода к другой сопутствующей информации (к другим статьям БЗ);
- значения слотов.

Ниже окна «Ответы» в строке ввода адреса документа можно прикрепить картинку (нажать на значок-ссылку «Послать фото» и указать путь к нему в открывшемся диалоге), вложенный документ («Послать файл» и нажать кнопку «Прикрепить»).

В случае необходимости вставить длинную ссылку можно использовать «Укорачивание ссылок».

Для этого вставить ссылку в строку укорачивателя и нажать «Укоротить». Ссылка должна вводиться в полном виде, с указанием протокола (например,

<https://yandex.ru/>). После чего нажать «Скопировать». После этого уже короткую ссылку можно вставлять в нужное место.

«Добавить кнопку» — это механизм добавления кнопки. «Кнопки» выводятся после текста ответа клиенту и служат для быстрого перехода клиента к другой информации, например, к другим статьям БЗ. Количество кнопок неограниченно.

Для создания новой кнопки требуется ввести идентификатор существующей (или будущей) статьи в БЗ и название, которое будет отображаться на самой кнопке (“Текст кнопки”).

Для удобства поиска записей в БЗ, на которые будут ссылаться кнопки, предусмотрена строка поиска. В нее достаточно ввести название записи. В случае успешного поиска будет выведена подсказка, на которую достаточно кликнуть мышкой, и поля идентификатор и текст кнопки заполнятся автоматически (рис. 3.15).

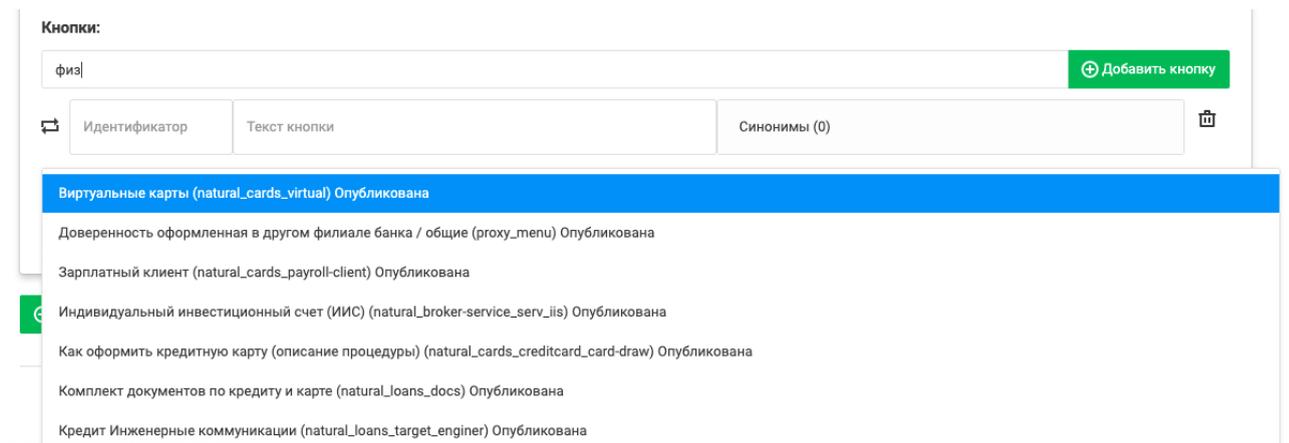


Рисунок 3.15

В случае необходимости кнопку можно удалить из ответа. Для удаления кнопки требуется нажать на красный крестик справа от строки с кнопкой и подтвердить удаление в открывшемся диалогом окне.

В случае необходимости ответ можно удалить из записи. Для удаления ответа надо нажать на кнопку удаления  справа от ответа и подтвердить удаление ответа в открывшемся диалоговом окне.

Статью можно деактивировать, т.е. исключить из выдачи и процесса обучения классификатора, нажав «Деактивировать», а также создать копию нажав «Создать копию статьи».

Платформа поддерживает подключение различных агентов для обработки запросов. Стандартные внутренние агенты:

- aiassist – чат-бот;
- routingagent - агент маршрутизации по операторам.

Дополнительно можно подключать внешних агентов (интеграции) - это, в том числе, позволяет для определенных навыков запускать внешние интеграции.

Например, `/redirect <agent_id>` Для переключения на агентов применяются три команды:

1. `/redirect` - передать текущий запрос для ответа другому агенту (т.е. разово вызвать агента и вернуться). Позволяет переслать текущее сообщение другому агенту.

2. `/switch` - меняет дефолтного агента для последующих сообщений, т.е. переключить дефолтного агента (агента, на которого поступит следующий запрос от клиента). Позволяет передать управление на другого агента и отключиться «рис.6.16».

3. `/switchredirect` - передать запрос и переключить дефолтного агента (комбинация 1+2 команд). Например, для перехода на какую-то статью в БЗ из текущего ответа в статье БЗ можно использовать агент aiassist и команду redirect.

Синтаксис у команды следующий:

```
/redirect aiassist [intent_id="INTENT_ID"]
```

Для перехода на статью с идентификатором "menu001" в поле ответа следует набрать следующий текст:

```
/redirect aiassist intent_id="menu001"
```

Ответы

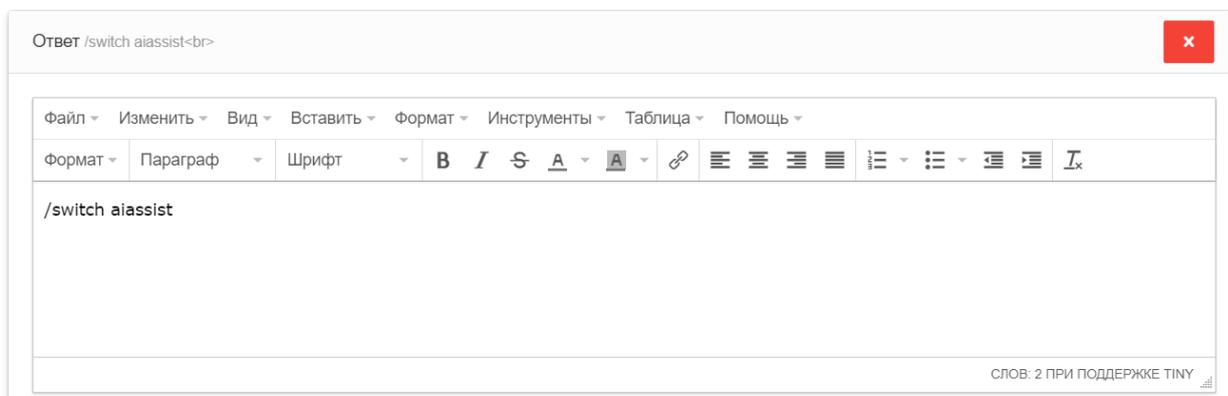


Рисунок 3.16

Следует иметь в виду, что в ответе, содержащем команду, не должно быть другой информации в поле текста ответа, а параметр "Тип текста ответа" обязательно должен быть проставлен как "Простой текст".

В статьях БЗ можно также использовать переменные, позволяющие получить данные, введенные пользователем в веб-чате при регистрации. Это переменные: `__FIRSTNAME__` и `__LASTNAME__` (используются только для первой версии aiassist).

Для обеспечения унифицированного форматирования сообщений в мессенджерах и социальных сетей должен поддерживаться формат форматирования сообщений markdown.

При использовании типа текста Markdown действуют следующие правила - Markdown работает в каналах Viber, Telegram, Vk, Facebook, Ok. В [Приложении 5](#) приведена информация по ограничениям markdown синтаксиса по всем каналам Платформы.

Поддерживаются следующие варианты форматирования сообщений:

1. Текст с форматированием:

Жирный текст. Пример: ****здесь текст****

Наклонный текст. Пример: *__здесь текст__*

Перечеркнутый текст. Пример: ~~~~здесь текст~~~~

2. Список.

Пример: элемент 1, элемент 2, элемент 3 .

3. Изображение.

Пример: ![Подпись изображения](http://image.com/image.png)

4. Блок кнопок.

Markdown делится на два блока, которые разделяются переводом строки \n:

Блок ответа.

Блок кнопок.

Правое окно автоматически заполняется примерным видом кнопок (оно может не совпадать с реальным отображением, из-за различающейся логики отображения и дополнительных ограничений в различных каналах).

Блок кнопок Markdown

Блок кнопок выделяется специальным синтаксисом:

Формат

```
```buttons(placement:inline fb_type:quick_reply)
...`
```

где

placement - тип отображения клавиатуры:

- inline - внутри сообщения (по-умолчанию),
- keyboard - внизу в клавиатуре.

fb\_type - тип отображения кнопок в Facebook Messenger, вариант по-умолчанию настраивается в настройках канала:

- quick\_reply - (по-умолчанию, если нет в настройках канала),
- button.

Внутри блока кнопок есть возможность размещать строки с кнопками с помощью знака « :: ». Каждая строка разделяется переводом строки \n и может содержать несколько кнопок.

```
```buttons
::
[Текст кнопки на первой строке](type:action action:test1 color:primary)
[Текст кнопки на первой строке](type:action action:test2
image:https://test.ru/test.png image_align:center image_emoji:⚡)
::
[Текст кнопки на второй строке](type:url action:https://google.com)
...`
```

Каждая кнопка имеет название заключенное в [], в названии кнопки можно использовать markdown синтаксис: [Текст ****кнопки**** в `~~markdown~~`], в каналах не поддерживающих markdown в кнопках, он будет вырезан.

Каждая кнопка имеет список атрибутов заключенных в () и для каждого атрибута имеется его название и через символ « : » его значение.

Атрибуты разделены пробелами, поэтому пробелы в названиях и значениях атрибутов запрещены.

На данный момент имеется поддержка следующих атрибутов:

type - тип кнопки, имеет фиксированный список значений:

- action - нажатие на кнопку обрабатывается как обычное нажатие с возвратом идентификатора или названия кнопки,

- url - нажатие на кнопку открывает ссылку, указанную в значении атрибута action,

- action - url ссылка или текстовый идентификатор кнопки, который будет возвращен при нажатии, зависит от параметра type.

visible - отображение кнопки

- true - включить отображение,

- false - отключить отображение,

- color - цвет кнопки, может быть одним из 3 типов (цвета типов - HEX, настраиваются в настройках канала).

secondary - обычная кнопка (по-умолчанию),

primary - кнопка подтверждение,

negative - кнопка отклонение,

image - картинка в кнопке (ссылка):

- image_align - положение картинки в кнопке,

- center - над текстом,

- left - слева от текста (по-умолчанию).

image_emoji - задать emoji символ, если канал не поддерживает картинки в кнопках:

- full_size - растягивание размера кнопки на весь размер inline клавиатуры в вайбере,

- true - применить растягивание,

- false - не растягивать (по-умолчанию).

Пример заполнения для создания кнопок представлен на рис. 3.17.

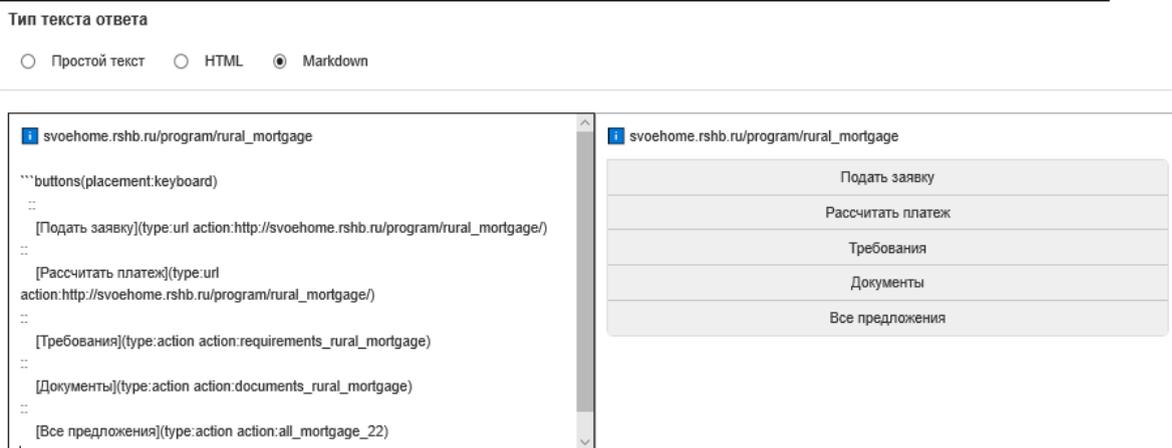
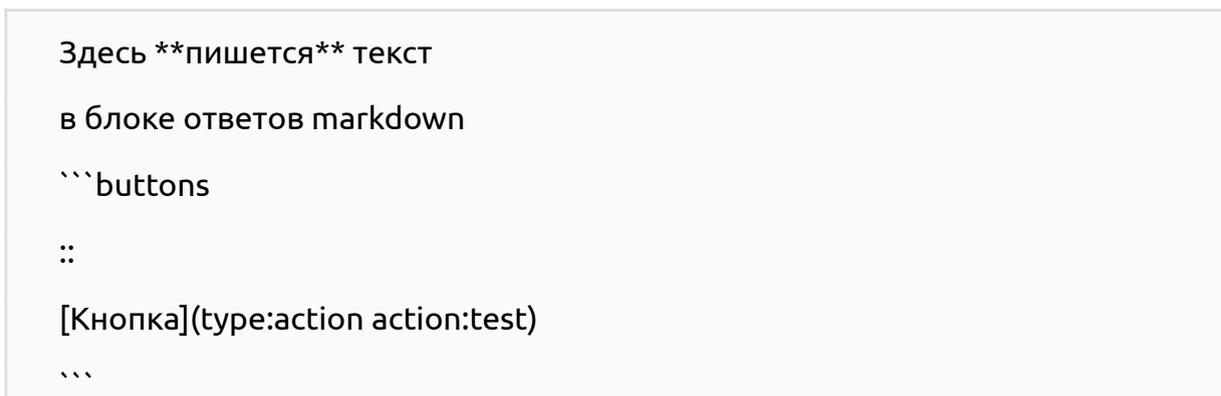


Рисунок 3.17

Блок ответов Markdown

Блок ответов находится строго перед блоком markdown и имеет вид:



В блоке ответов разрешено использование стандартных тэгов markdown.

- Жирный текст. Пример: **здесь текст**.
- Наклонный текст. Пример: *здесь текст*.
- Перечеркнутый текст. Пример: ~~здесь текст~~.
- Список.

Пример: **элемент 1**, *элемент 2*, ~~элемент 3~~.

Ниже приведен пример заполнения для создания кнопок с блоком текста, результат представлен на рис. 3.18.

```

```buttons
some text _some text_
__some text__ some text

~~some text~~

```buttons
::
[**Кнопка1**](type:url action:http://google.com)
[_Кнопка2_](type:action action:test2)
::
[_Кнопка3_](type:url action:http://google.com)
[~~Кнопка4~~](type:action action:test4)

```

Тип текста ответа

Простой текст
 HTML
 Markdown

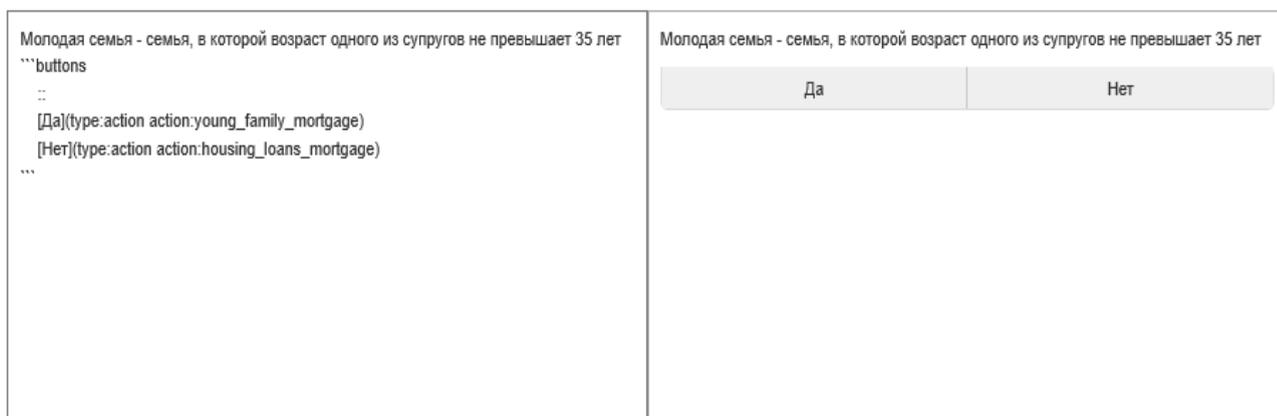


Рисунок 3.18

В блоке ответов возможно использование изображений markdown вида:

```
![Подпись изображения](http://image.com/image.png).
```

В случае наличия изображений в блоке ответа, в канал будет отправлено несколько ответов: с текстом перед изображением, с изображением, с текстом (и

кнопками) после изображения. Подпись к изображению будет отправлена в каналах, которые ее поддерживают. Изображений может быть несколько. Лучшая практика не делать изображение последним в блоке ответов, т.к. в таком случае блок кнопок не отправится в канал без блока ответов. Большинство каналов не поддерживают отправку только кнопок без текста.

Пример:

Текст ответа перед изображением.

```
![Подпись изображения](http://image.com/image.png)
```

Текст ответа после изображения. Этот текст будет отправлен вместе с кнопками ниже.

```
```buttons
::
[**Кнопка1**](type:url action:http://google.com)
[_Кнопка2_](type:action action:test2)

::
[_Кнопка3_](type:url action:http://google.com)
[~~Кнопка4~~](type:action action:test4)
```
```

Стикер в markdown

В блоке ответов возможно использование стикера:

```
```sticker(id:12345)
```
```

id - идентификатор стикера (может быть из конкретного канала или общий в системе).

Поддержка отправки в каналы: Вайбер, Telegram, Одноклассники, Фейсбук (как картинка), Вконтакте (как картинка).

Прием стикеров Системой КТ

В данный момент поддерживается прием стикеров из следующих мессенджеров:

- Facebook,
- Odnoklassniki,
- Telegram,
- Viber,
- VKontakte.

В данный момент поддерживается отправка стикеров из Системы КТ в следующие мессенджеры:

- Facebook (отправка картинкой),
- Odnoklassniki,
- Telegram,
- Viber,
- VKontakte (отправка картинкой).

Для того, чтобы бот мог отправлять стикеры, необходимо сначала завести стикеры в Систему КТ (см. п. [Стикеры](#)).

Разделитель текстовых ответов markdown

В блоке ответов возможно установить разделитель между текстовыми ответами:

Первый ответ отдельным сообщением

```
```nextaswer
```

```
```
```

Второй ответ отдельным сообщением

3.6 Раздел «Слоты»

Слоты – это информация, которая хранится в контексте диалога и может быть запрошена у Обратившегося, либо получена другим образом в процессе обработки диалога.

Слоты могут использоваться для формирования различных ответов в зависимости от значений слотов. Это достигается путём добавления к ответу условий, при которых ответ актуален. Таким образом, с помощью слотов можно формировать диалог с пользователем с запросом нужной информации, которую в дальнейшем можно использовать для формирования ответа и, в том числе, передавать на внешние интеграции.

Слоты также можно использовать для формирования опросов и анкеты для оператора.

Опросы пользователей могут также конфигурироваться в Базе знаний и отправляться пользователю ИИ Ботом или Оператором.

Таким образом, с помощью слотов можно формировать диалог с Обратившимся с запросом нужной информации, которую в дальнейшем можно использовать для формирования ответа и, в том числе, передавать на внешние интеграции.

Слоты в статье БЗ добавляются на вкладку "Слоты" в том порядке, в котором будет запрашиваться информация из них, согласно сценарию. Слоты под ответом добавляются в том случае, если от их значения, выбранного/введенного пользователем в процессе диалога, зависит ответ статьи.

Ответ может зависеть от значения слота, если он задействован в ответе.

Для создания слота необходимо перейти в раздел «Слоты» и нажать кнопку «Добавить» (рис. 3.19).

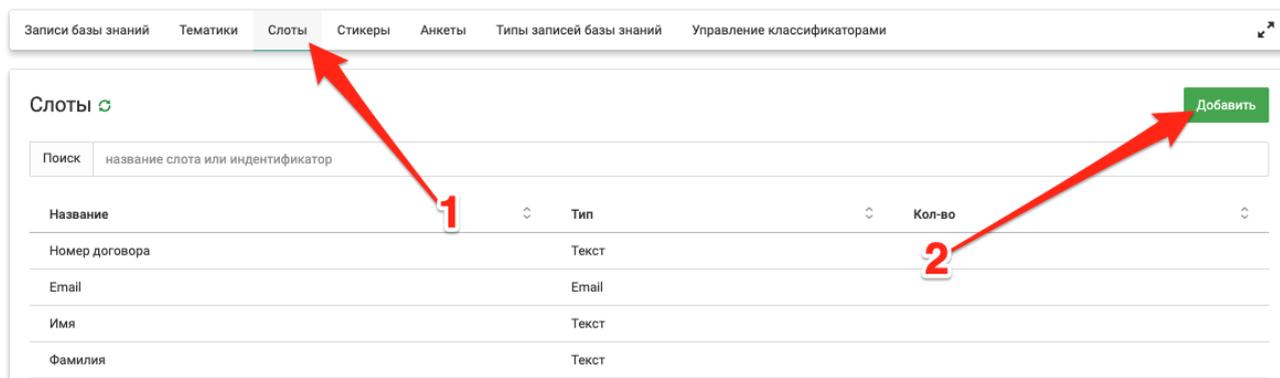


Рисунок 3.19

После этого необходимо заполнить все поля всплывающей формы (рис. 3.20).

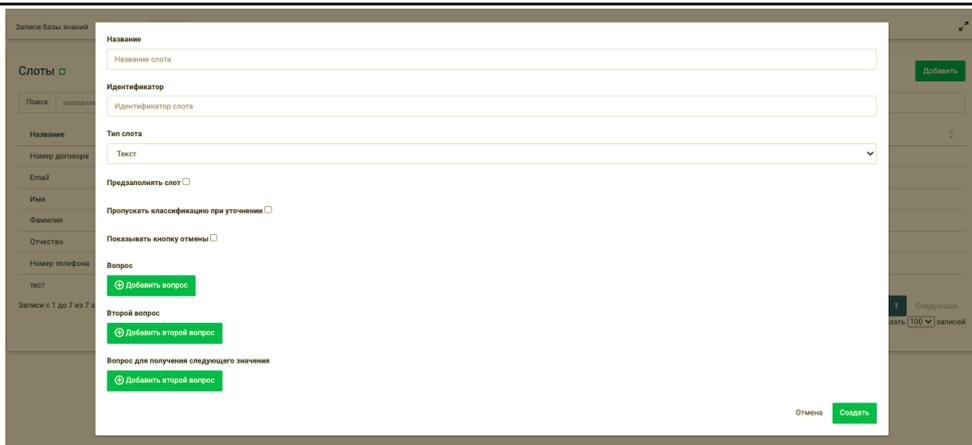


Рисунок 3.20

Если при заполнении слота выбрать тип «Вложения» и «Регулярное выражение» (рис. 3.21), появятся дополнительная строка «Регулярное выражение» (рис. 3.22).



Рисунок 3.21

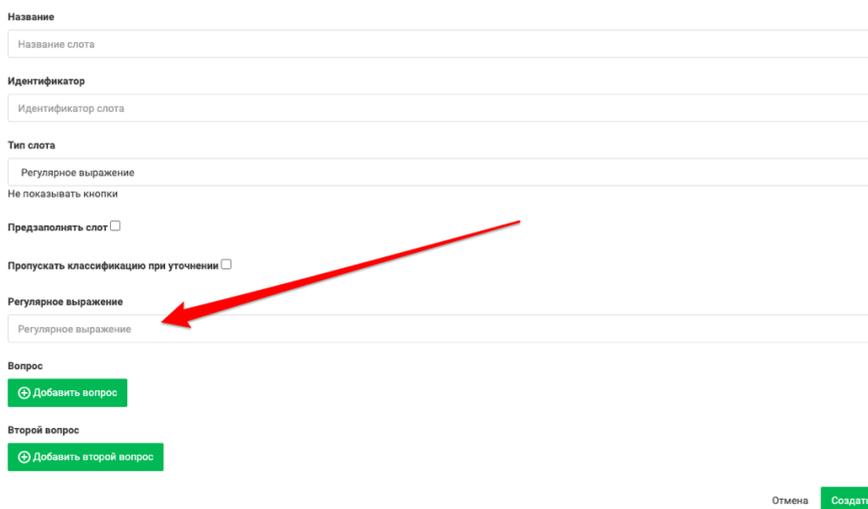
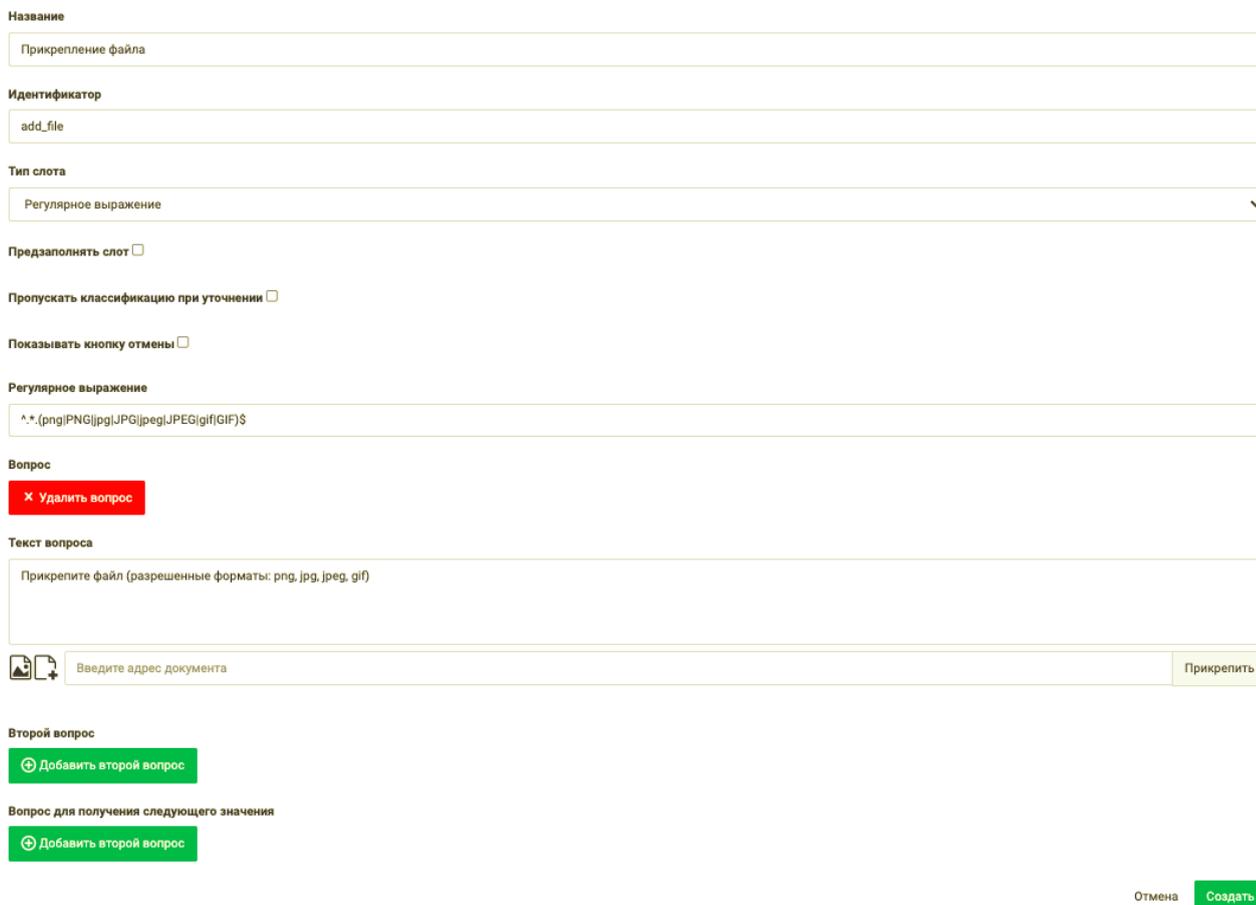


Рисунок 3.22

На рисунке 5.23 приведен пример заполнения информации слота «вложение».



Название
Прикрепление файла

Идентификатор
add_file

Тип слота
Регулярное выражение

Предзаполнять слот

Пропускать классификацию при уточнении

Показывать кнопку отмены

Регулярное выражение
^.*(png|PNG|jpg|JPG|jpeg|JPEG|gif|GIF)\$

Вопрос
X Удалить вопрос

Текст вопроса
Прикрепите файл (разрешенные форматы: png, jpg, jpeg, gif)

Введите адрес документа Прикрепить

Второй вопрос
+ Добавить второй вопрос

Вопрос для получения следующего значения
+ Добавить второй вопрос

Отмена

Рисунок 3.23

Слот содержит (см. рис. 3.23):

- Название слота, отображаемое в таблице слотов и в выпадающих списках в форме создания новых записей в Базе знаний.
- Идентификатор слота — обязательное к заполнению поле. Допускается использование строчных букв латинского алфавита, цифры, отдельных символов (тире, нижнее подчеркивание) и их комбинаций без пробелов.
- Тип слота (выбирается из выпадающего списка).
- Предзаполнение слота - Слоты с этой галкой пытаются заполниться на каждый ввод пользователя, пока статья в БЗ (интент) не выдаст ответ. Без галки слот может заполниться только с помощью уточнений бота.

- Добавить вопрос — этот вопрос задается, а вложения посылаются Обратившемуся в диалоге (в начале работы слота). Сообщение о формате ввода и т.п.

- Добавить второй вопрос — это вопрос-уточнение, если ответ на первое сообщение был неправильным или надо продолжить диалог. Здесь может быть дополнительно сообщение, например, о том что неправильно введена информация, например, неправильный формат файла.

- Добавить второй вопрос (Вопрос для получения следующего значения) — заполняется в случае, если предполагается прикрепление вложений более 1 раза и установлены галочки «Пропускать классификацию при уточнении» и «Не показывать кнопки для словаря».

Установка галочки в Предзаполнение необходимо для того, чтобы попытаться заполнить слот из вопроса клиента.

Например, у клиента спрашивают номер из паспортных данных, а он в ответ на этот запрос вместе с номером еще и дату выдачи напишет. А это был второй слот (т.е. следующий вопрос по сути, который хотели спросить после номера паспорта), а т.к. слот «Дата выдачи паспорта» - это дата или регулярное выражение на дату, то он может автоматически заполниться и уже не спрашиваться у клиента.

Таким же образом можно пытаться из контекста вопроса заполнить слот типа словарь, например.

Устанавливать галочку на предзаполнение слота типа текст не рекомендуется, т.к. текст - это все, что угодно, и этот слот всегда будет заполнен, в том числе бессмыслицей.

В Системе КТ доступны следующие типы слотов:

1. Регулярное выражение — значением слота будет значение, полученное из ответа Обратившегося при помощи регулярного выражения.

Полное описание синтаксиса регулярного выражения можно найти по ссылке - <https://docs.python.org/3.8/library/re.html#regular-expression-syntax>.

Проверить правильность заполненного формата можно по следующей ссылке - <https://regex101.com/>.

2. Текст — значением слота будет произвольный текст (заполняется любым вводом, показывает кнопку "Отменить").

3. Целое число — значением слота будет целое число.

4. Словарь — это ограниченный список значений (значением слота является id значения словаря, а не текст кнопки). При его выборе открывается механизм заполнения значений словаря, в котором следует нажать кнопку «Добавить значение» и ввести в открывшиеся поля уникальный идентификатор, с которым будет работать система, и значение, видимое пользователям (в виде надписей на кнопках).

5. Дата - значением слота будет дата (формат даты: 09.11.2020, 09 11 2020, 09-11-2020, 09/11/2020, 9 11 20).

6. Время - значением слота будет время (формат времени: 15:30).

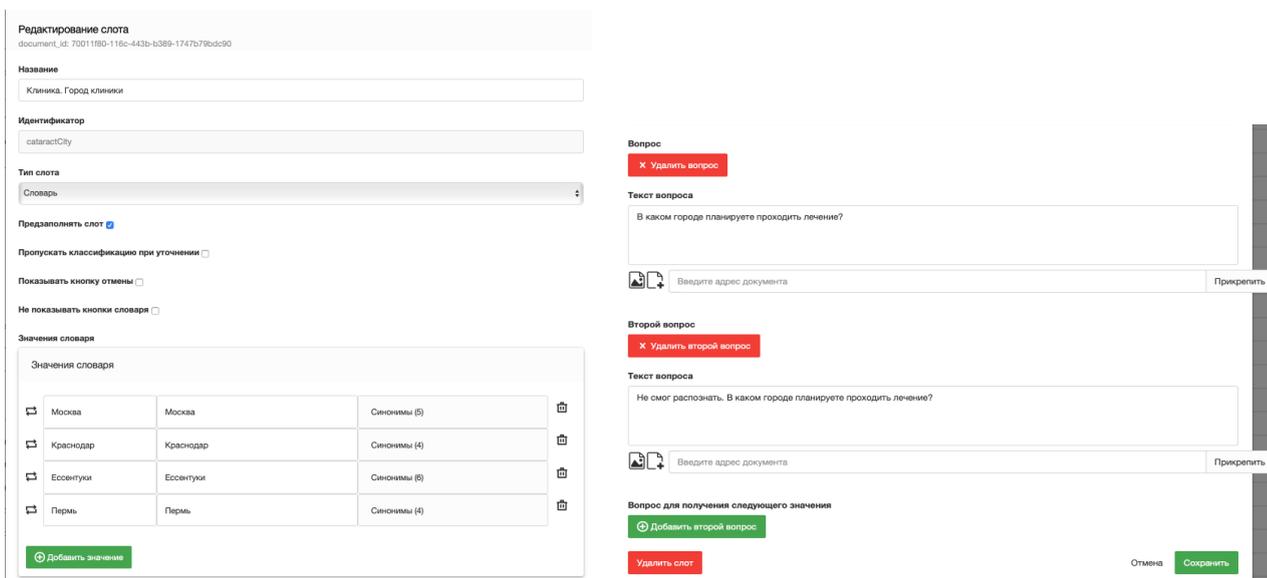
7. Вложения – в качестве вложения может быть любой файл. С каким расширением (и/или именем и расширением) будут приниматься файлы, описывается с помощью регулярного выражения, которое добавляется в настройках слота.

8. Email - значением слота будет e-mail.

9. Внешний словарь (Использует варианты ответов не из настроек слота, а из слота slot_id_OPTIONS)

Для регулярных выражений заполняется соответствующее поле формы «Регулярное выражение».

Пример слота для вопроса о городе, где пользователь хочет обратиться в клинику (с 4 возможными вариантами ответов), приведен на рисунке 5.24.



Редактирование слота
document_id: 7001180-116c-443b-b389-1747b79bdc90

Название
Клиника. Город клиники

Идентификатор
catarsactCity

Тип слота
Словарь

Предзаполнить слот

Пропускать классификацию при уточнении

Показывать кнопку отмены

Не показывать кнопки словаря

Значения словаря

| Значения словаря | | |
|------------------|-----------|--------------|
| Москва | Москва | Синонимы (5) |
| Краснодар | Краснодар | Синонимы (4) |
| Ессентуки | Ессентуки | Синонимы (6) |
| Пермь | Пермь | Синонимы (4) |

Вопрос

Текст вопроса
В каком городе планируете проходить лечение?

Второй вопрос

Текст вопроса
Не смог распознать. В каком городе планируете проходить лечение?

Вопрос для получения следующего значения

Рисунок 3.24

Разберем следующую ситуацию. Необходимо добавить в Базу знаний информацию о смене тарифа (вопрос «Как я могу сменить тариф?»). Ответ на этот вопрос зависит от типа лица, заключившего договор (физическое или юридическое лицо). Для физического лица должна быть выведена инструкция №1, для юридического – с текстом №2. Также для юридического лица необходимо сохранить подсказку в БЗ для оператора (если ответ на вопрос будет обрабатываться не автоматически ботом), которая не должна быть видна клиенту.

На первом шаге создадим слот с названием «Определение типа клиента» (меню «База знаний» -> вкладка «Слоты»). Заполним поля формы создания слота (рис. 3.25).

Название
Определение типа клиента

Идентификатор
b2b_b2c_client

Тип слота
Словарь

Предзаполнять слот

Пропускать классификацию при уточнении

Показывать кнопку отмены

Не показывать кнопки словаря

Значения словаря
Значения словаря

Вопрос
X Удалить вопрос

Текст вопроса
Вы являетесь юридическим или физическим лицом?

Введите адрес документа Прикрепить

Второй вопрос
+ Добавить второй вопрос

Вопрос для получения следующего значения
+ Добавить второй вопрос

Отмена Создать

Рисунок 3.25

Так как мы выбрали тип слота «Словарь», то необходимо внести значения в словарь (рис. 3.26).

Значения словаря

Значения словаря

| | | | | |
|---|------------|------------------|--------------|---|
|  | b2b_client | Юридическое лицо | Синонимы (0) |  |
|  | b2c_client | Физическое лицо | Синонимы (0) |  |

+ Добавить значение

Рисунок 3.26

Перейдем на вкладку «Записи базы знаний» и создадим новую статью с именем «Изменение тарифного плана» (рис. 3.27 – 5.28).

Название

Идентификатор

Тип записи

Справочник
 Ответы на вопросы пользователя
 Общие вопросы

Настройки

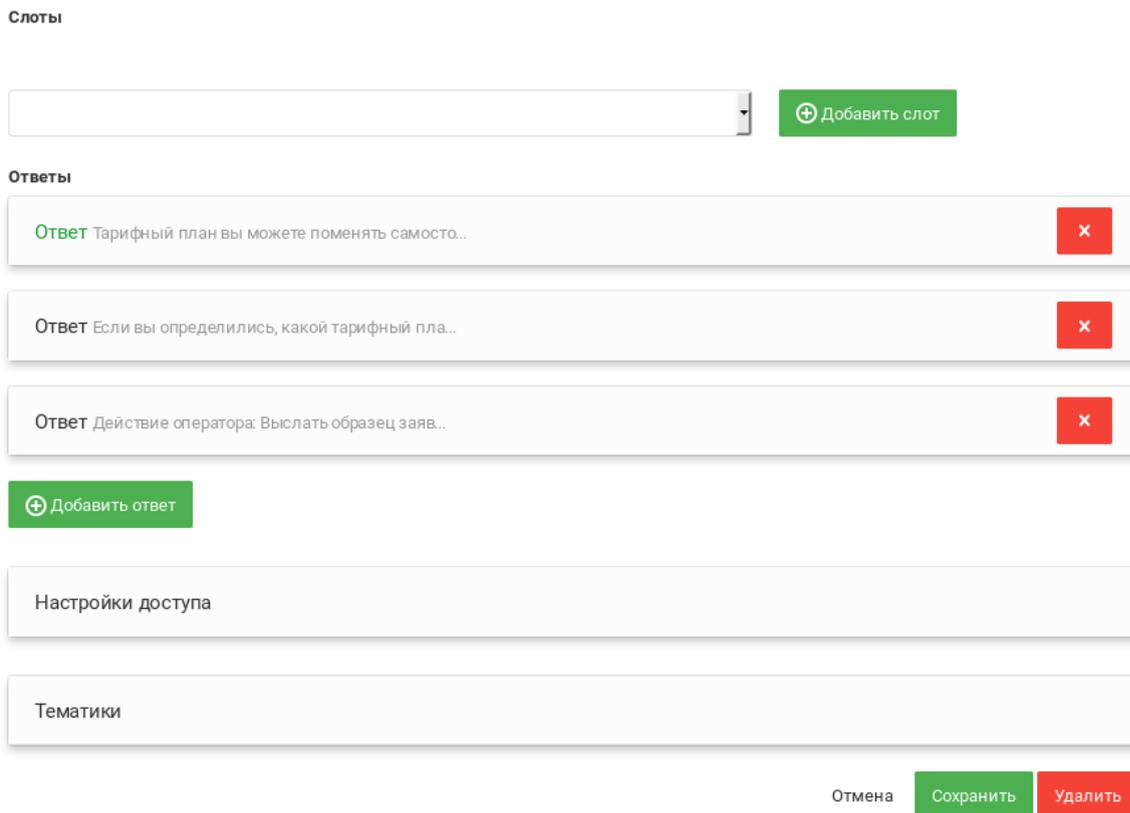
Политика отправки сообщений

Все подходящие
 Первое подходящее
 Случайное подходящее

Когда пользователь спрашивает:

| | |
|------------------------------------|---|
| Новый вопрос |  |
| Как я могу поменять тарифный план? |  |
| как сменить тарифный план |  |
| замена тарифного плана |  |
| смена тарифа |  |
| изменить тарифный план |  |
| Изменение тарифного плана |  |

Рисунок 3.27



The screenshot shows a user interface for configuring a slot and its answers. At the top, under the heading "Слоты" (Slots), there is a dropdown menu and a green button labeled "Добавить слот" (Add slot). Below this, under the heading "Ответы" (Answers), there are three answer cards. Each card starts with "Ответ" (Answer) followed by a text snippet and a red "X" delete button. The answers are: "Тарифный план вы можете поменять самосто...", "Если вы определились, какой тарифный пла...", and "Действие оператора: Выслать образец заяв...". Below the answers is a green button "Добавить ответ" (Add answer). At the bottom of the interface, there are three buttons: "Отмена" (Cancel), "Сохранить" (Save), and "Удалить" (Delete). Two additional input fields are visible: "Настройки доступа" (Access settings) and "Тематики" (Topics).

Рисунок 3.28

Для возможности использования слота внутри записи БЗ необходимо выбрать в выпадающем списке «Слоты» все названия слотов, которые будут использоваться в создаваемой записи БЗ.

В данной записи необходимо создать три ответа.

Первый ответ для физического лица (с использованием значения «Физическое лицо» из словаря слота «Определение типа клиента») (рис. 3.29). Второй ответ для юридического лица (с использованием значения «Юридическое лицо» из словаря слота «Определение типа клиента»). Третий ответ – это подсказка оператору действий при работе с юридическим лицом (с использованием значения «Юридическое лицо» из словаря слота «Определение типа клиента»).

Ответ Тарифный план вы можете менять самосто... ✕

Текст ответа

Тарифный план вы можете менять самостоятельно в своей Личном кабинете.
Слева на главной странице в меню «Тарифы и услуги» выберите пункт «Смена услуг». Далее выберите строку с текущим тарифом. Далее выбираете опции, необходимые вам, и нажимаете «Далее». Смена тарифного плана произойдет по окончании текущего расчетного периода.
По желанию можно выбрать отметкой «Служба смены услуг». Смена произойдет в течение нескольких минут.

Вложения:

Прикрепить: Прикрепить

Кнопки:

⊕ Добавить кнопку

Значения слотов

⊕ Добавить слот

Слот Определение типа клиента ✕

✕

⊕

Рисунок 3.29

При создании ответа в выпадающем списке «Значения слотов» выбирается слот, который используется в текущем ответе. Далее, при нажатии на имя выбранного слота в выпадающем списке, выбирается значение, при котором будет выбран данный ответ.

Чтобы юридическому лицу был виден только один ответ, предназначенный для клиента, была выбрана политика отправки сообщений «Первое подходящее».

В строке «Слоты» можно создать привязку новой записи к одному или нескольким из существующих слотов. Для этого нажать треугольник и выбрать слот. Для привязки к другому слоту нажать «Добавить слот» и выбрать ещё одну привязку.

После заполнения всех разделов создания новой записи Базы знаний, проходим стандартную процедуру публикации статьи (подробно см. п.6.9). Для этого создаем черновик, затем отправляем на проверку, после чего публикуем статью. Далее, необходимо обучить бота в разделе “управление классификаторами”.

Теперь при вопросе пользователя “Как я могу сменить тарифный план?” ему будет сначала задан вопрос от том, физическим или юридическим лицом является данный клиент, и в зависимости от его ответа будет выдано сообщение для данного типа клиента (рис. 3.30).

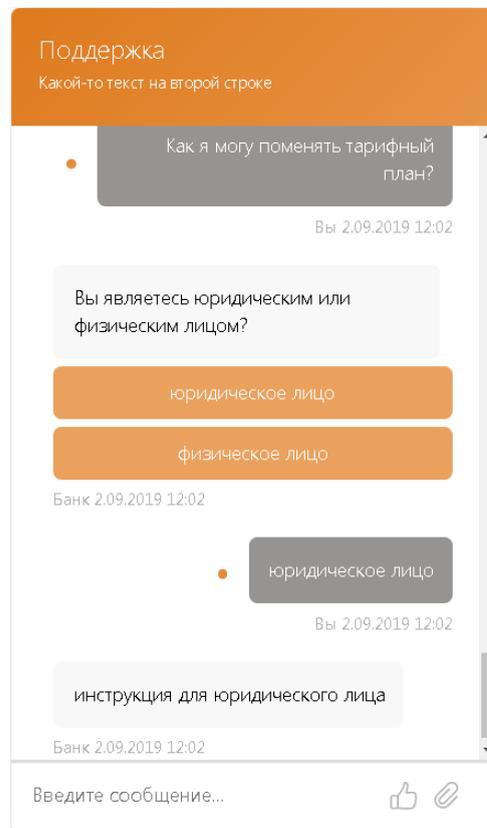


Рисунок 3.30

Разберём ещё одну ситуацию – Создание слота «Каналы», возможность добавления ответов для различных каналов в рамках одной статьи.

Разделение ответов по каналам возможно настроить в том случае, если для каждого из ответов в рамках одной статьи будет добавлен слот “Каналы”.

Слот создается один раз и далее используется в статьях при необходимости разделения ответов в рамках одной статьи по каналам.

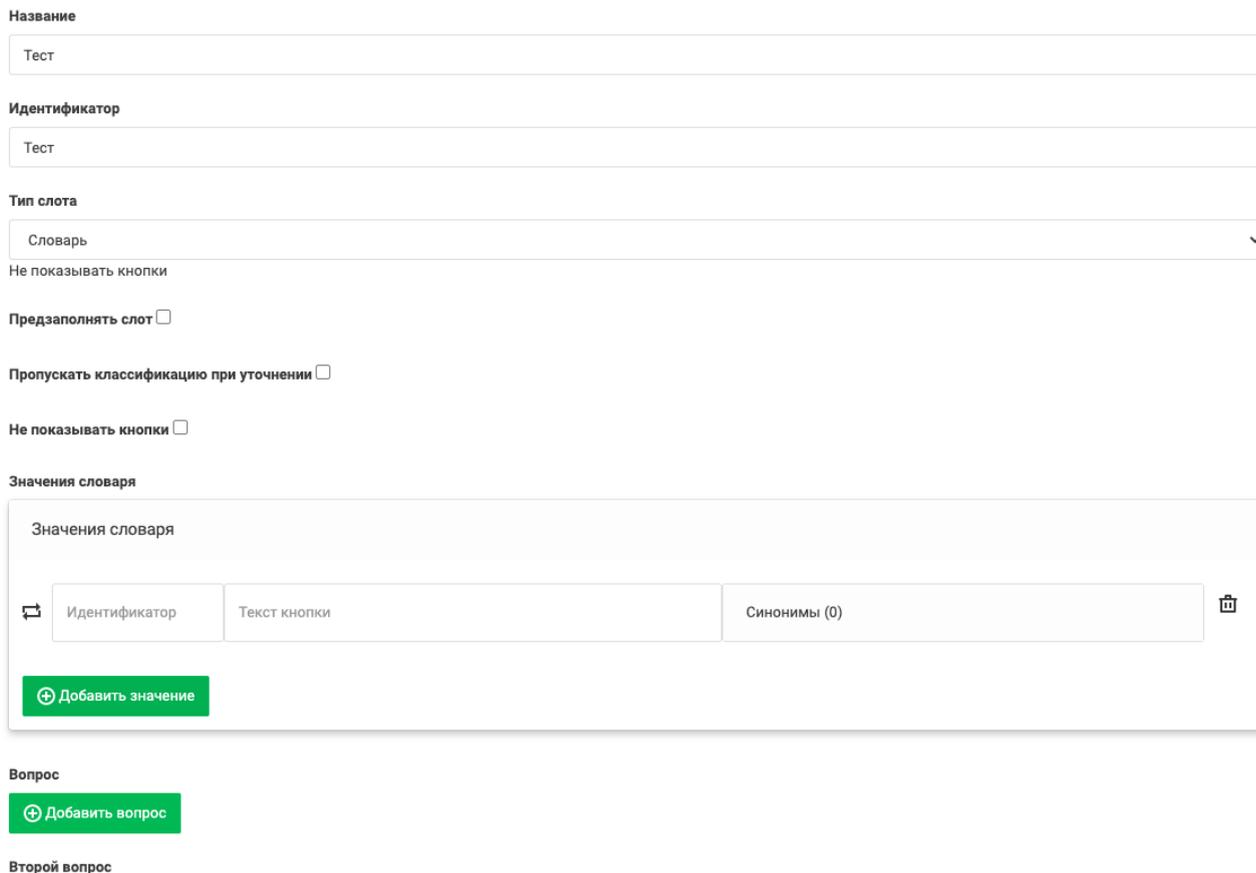
В меню «База знаний» -> вкладка «Слоты» нажать кнопку “Добавить” и заполнить поля открывшейся формы:

- поле «Название»: Каналы;
- поле «Идентификатор»: channel_id;

Важно! Менять название идентификатора нельзя, он обязательно должен быть таким!

- поле «Тип слота»: Словарь (выбрать из выпадающего списка).

Далее необходимо нажать на панель «Значения словаря», далее нажать на кнопку «Добавить значение» и заполнить поля «Идентификатор» и «Название» для каждого из каналов, которые применяются в системе (в случае, если каналов более одного, нажать на кнопку «Добавить значение» повторить) (рис. 3.31).



Название

Тест

Идентификатор

Тест

Тип слота

Словарь

Не показывать кнопки

Предзаполнять слот

Пропускать классификацию при уточнении

Не показывать кнопки

Значения словаря

Значения словаря

| Идентификатор | Текст кнопки | Синонимы (0) |
|---------------|--------------|--------------|
|---------------|--------------|--------------|

Добавить значение

Вопрос

Добавить вопрос

Второй вопрос

Рисунок 3.31

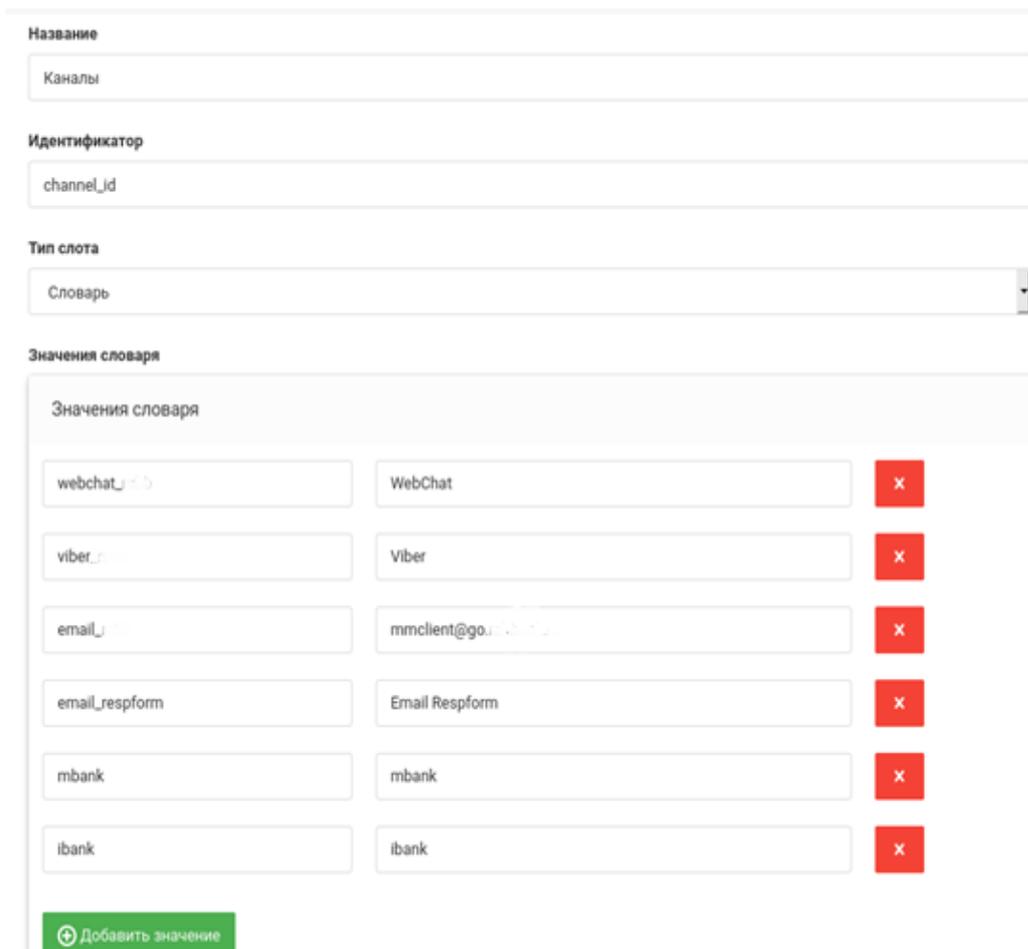
Идентификаторы значений словаря должны соответствовать идентификаторам каналов, прописанных в меню "Настройки".

Пример идентификаторов каналов:

- webchat (WebChat);
- viber (Viber);
- email;
- email_respform (Email Respform).

Важно! Вы можете изменять названия каналов, которые будут отображаться в выпадающих списках при последующем использовании слота Каналы, но идентификатор слота и идентификаторы значения словаря изменению не подлежат, они обязательно должны быть занесены в слот-словарь Каналы без изменений.

После заполнения всех полей нажать кнопку «Создать» внизу формы (рис. 3.32).



| Значения словаря | | |
|------------------|----------------|---|
| webchat_ | WebChat | X |
| viber_ | Viber | X |
| email_ | mmclient@go... | X |
| email_respform | Email Respform | X |
| mbank | mbank | X |
| ibank | ibank | X |

⊕ Добавить значение

Рисунок 3.32

Правила создания нескольких ответов для нескольких каналов в рамках одной статьи описаны ниже в описании меню «База знаний» вкладка «Записи Базы знаний» - (см. [Записи базы знаний](#)).

При редактировании записи можно отредактировать слот. При этом возможно дополнительно определить привязку записи к слотам либо удалить существующие привязки. На этой же вкладке можно изменить порядок следования

слотов (в каком порядке они будут выводиться для запроса клиенту). Для этого достаточно просто перетащить их, удерживая мышкой (рис. 3.33).

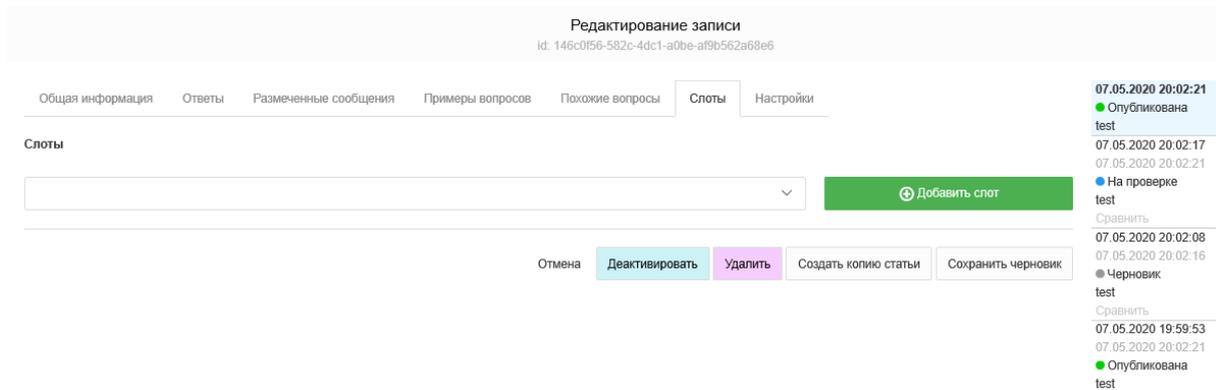


Рисунок 3.33

Пример работы с товарами в ВКонтакте, Facebook Messenger, Одноклассники.

При поступлении обращения с приложенным товаром из социальной сети, платформа CraftTalk обеспечивает возможность распознавания товара ботом. Для этого платформа заполняет следующие слоты:

- sku_id - идентификатор товара;
- sku_title - название товара;
- has_sku - признак наличия.

Далее можно стандартными средствами платформы обеспечить реакцию бота на такой запрос.

3.6.1 Пример создания статьи для распознавания товара в социальной сети ВКонтакте

1. Добавляем слот sku_id, если он не создан (рис. 3.34).

Редактирование слота

Название

Идентификатор

Тип слота

Текст

Вопрос

⊕ Добавить вопрос

Вопрос при заполненном слоте

⊕ Добавить вопрос при заполненном слоте

Удалить слот

Отмена Сохранить

Рисунок 3.34

2. Добавляем статью с примерами вопроса “Здравствуйте! Меня заинтересовал этот товар.” – такой вопрос передает Вконтакте при отправке сообщения с товаром. Это позволит обращению с приложенным товаром попасть на нужную статью базы знаний.

3. Из ссылки на товар извлекаем идентификатор товара. Например, для товара на рисунке 5.35 это – 3530481.

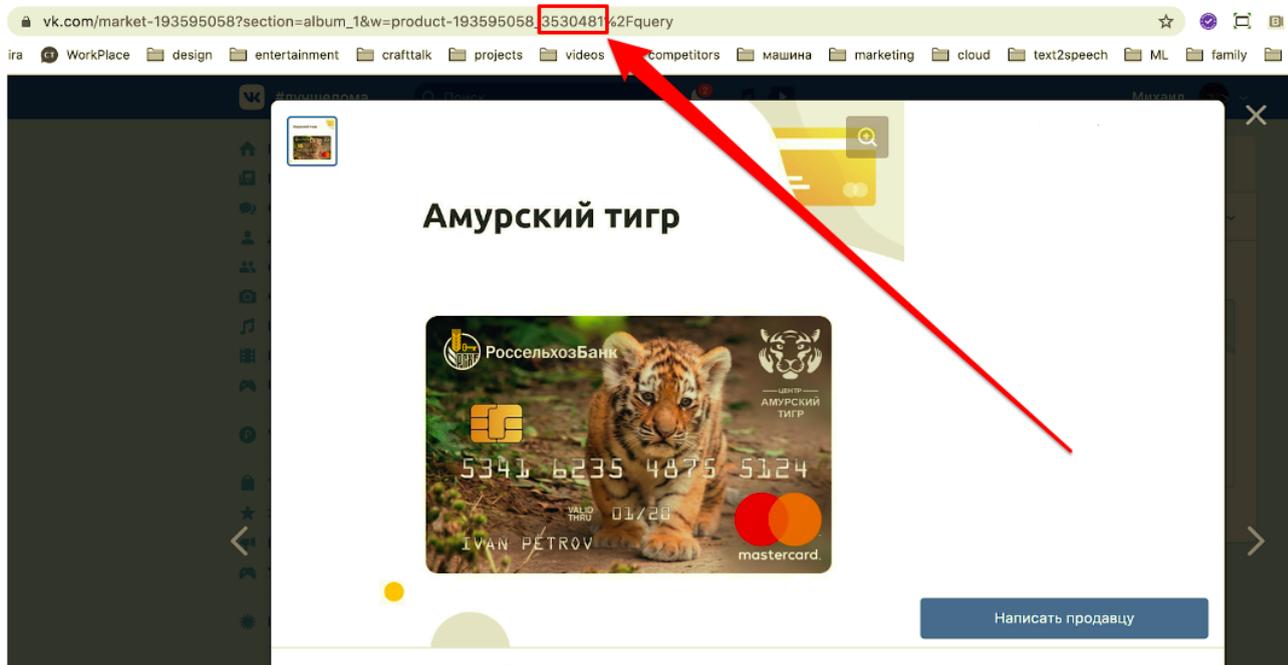


Рисунок 3.35

4. В статье добавляем ответ с ограничением на слот "id товара" (рис. 3.36). Этот ответ будет выдаваться для товара с указанным идентификатором.

Значения слотов:



Рисунок 3.36

5. В статье добавляем ответ без ограничений по значениям слотов – данный ответ будет выдаваться когда товара нет или он не распознан другими ответами.

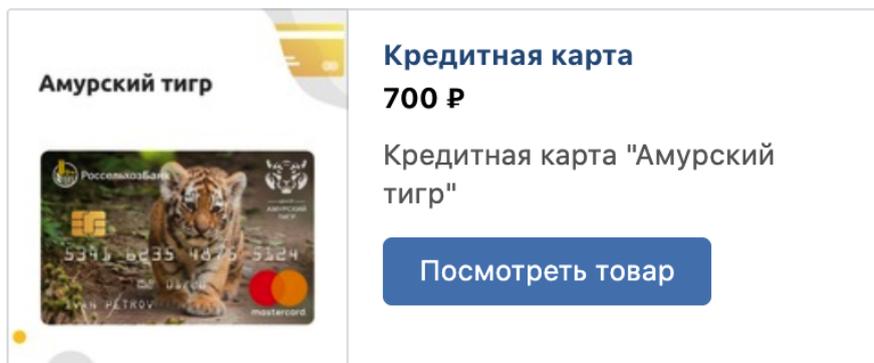
6. Результат настройки Базы знаний на распознавание товара Вконтакте приведен на рисунке 5.37.



Михаил 13:26

Здравствуйте!

Меня заинтересовал этот товар.



13:26

Спасибо, что заинтересовались Амурским тигром!

Рисунок 3.37

3.7 Раздел «Настройки» (создание статьи БЗ - настройки)

Раздел «Настройки» позволяет управлять выдачей сообщений.

Для «Политика отправки сообщений» возможны следующие варианты «рис..3»:

1. Все подходящие – будут выведены все ответы, подходящие по заданным в ответах значениям слотов.

2. Первое подходящее будет выведен один, максимально удовлетворяющий запросу, ответ. Это наиболее часто применяемый тип политики сообщений.

3. Случайное подходящее – будет выведен случайный ответ из всех найденных подходящих (например, одно из возможных приветствий Обратившегося).

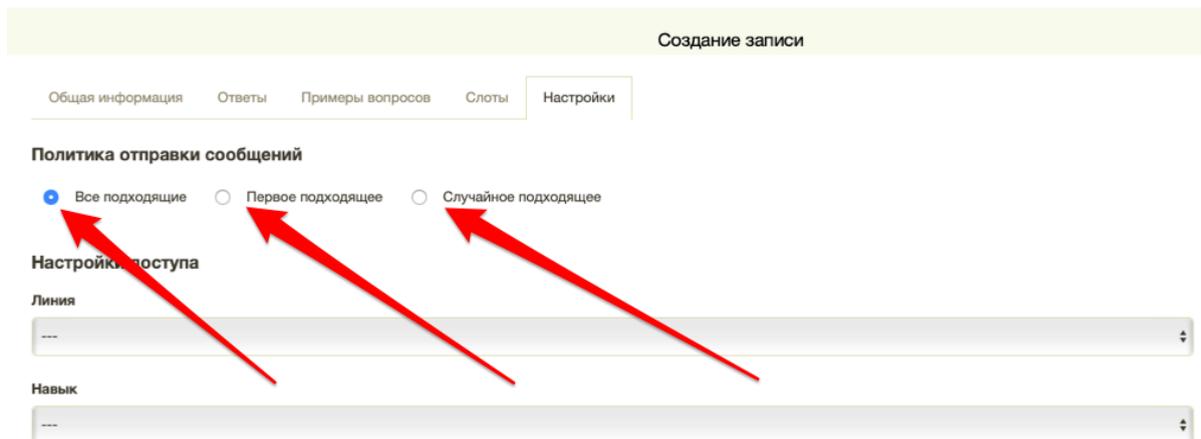


Рисунок 3.38

В «Настройки доступа» можно отрегулировать доступ к данной статье со стороны Операторов (рис. 3.39).

Настраиваются:

- Линия;
- Навык;
- Группа;
- Язык.

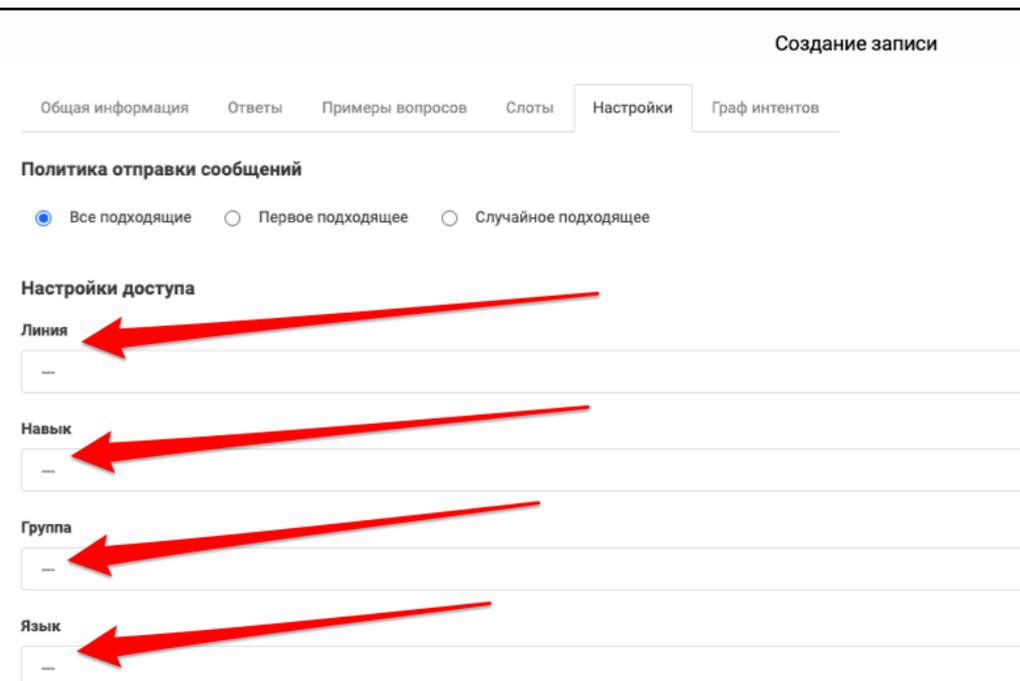


Рисунок 3.39

Настройка осуществляется одинаково нажатием уголка справа в строке и выбором требуемой настройки.

Сохранение записи осуществляется процедурой – Сохранить черновик > отправить на проверку > опубликовать.

Так же, при выходе нажатием «отменить» в появившемся уведомлении есть возможность сохранить статью БЗ (рис. 3.40).

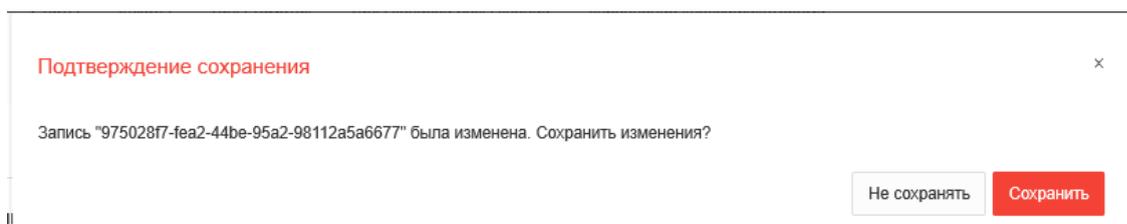


Рисунок 3.40

Изменение будет сохранено и статью БЗ можно в дальнейшем опубликовать.

3.8 Пример создания нескольких ответов для нескольких каналов в рамках одной статьи

Для того чтобы настроить несколько ответов для нескольких каналов в рамках одной статьи необходимо:

1. Зайти в пункт меню «База знаний», вкладка «Записи базы знаний».
2. Открыть для редактирования статью, в которую следует добавить ответ для другого канала или создать новую статью, если она еще не создана в Базе знаний.
3. Перейти на вкладку «Слоты» в форме редактирования / создания статьи Базы знаний и выбрать из выпадающего списка слот «Каналы» (рис. 3.41).

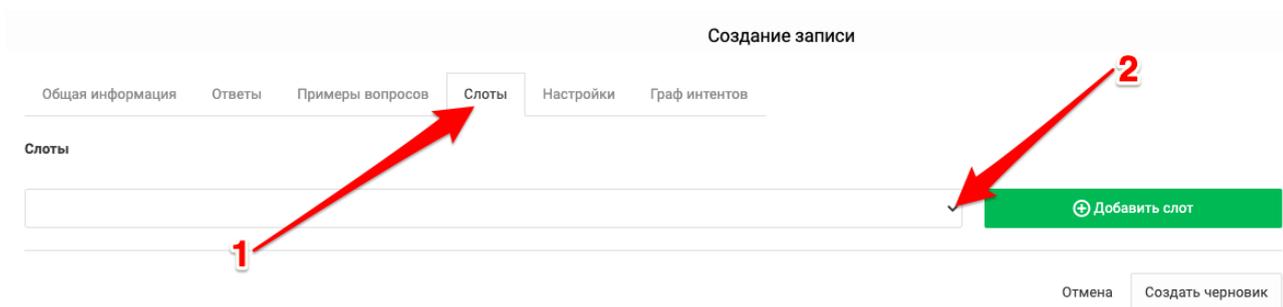


Рисунок 3.41

4. Перейти на вкладку «Ответы».
5. Создать ответ, который должен уходить в один из каналов, например, в канал WebChat (кнопка «Добавить ответ»). Если ответ уже создан в рамках статьи, то ответ создавать не требуется, просто пропустить этот этап).
6. Добавьте слот Каналы под ответом для вебчата - выбрать из выпадающего списка «Значения слотов» пункт «Каналы» и нажать «Добавить слот» (рис. 3.42 - 5.43).

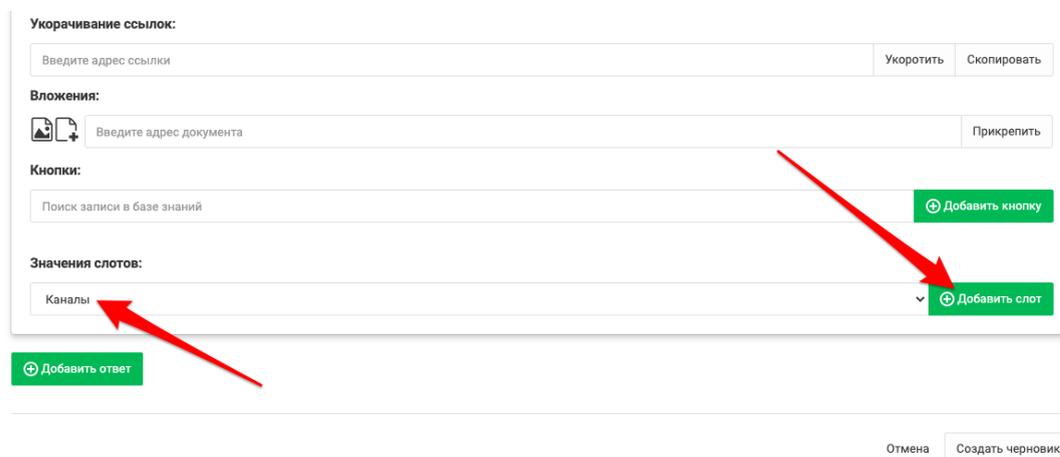


Рисунок 3.42

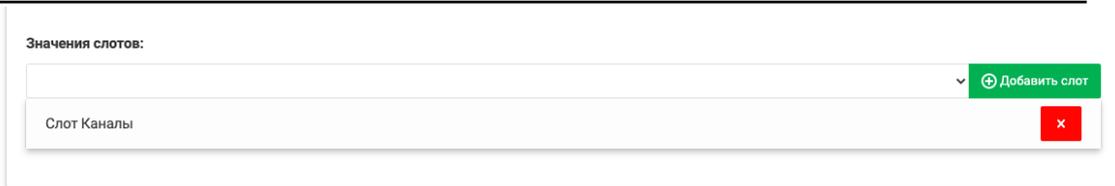


Рисунок 3.43

7. Раскрыть панель «Слот Каналы» и нажать на зеленую кнопку «+» (рис. 3.44).

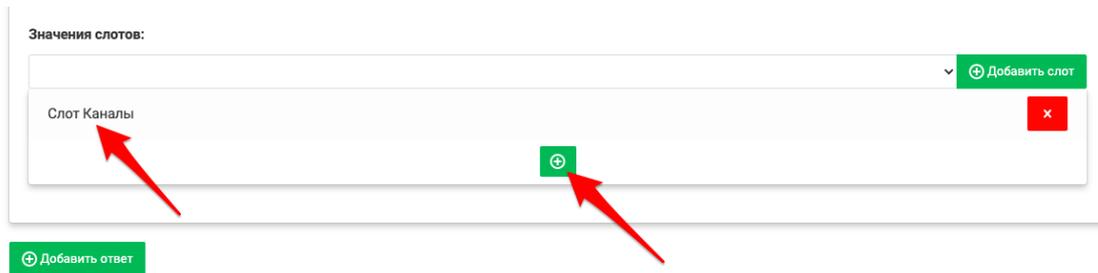


Рисунок 3.44

8. В появившемся выпадающем списке выбрать канал WebChat (рис. 3.45).

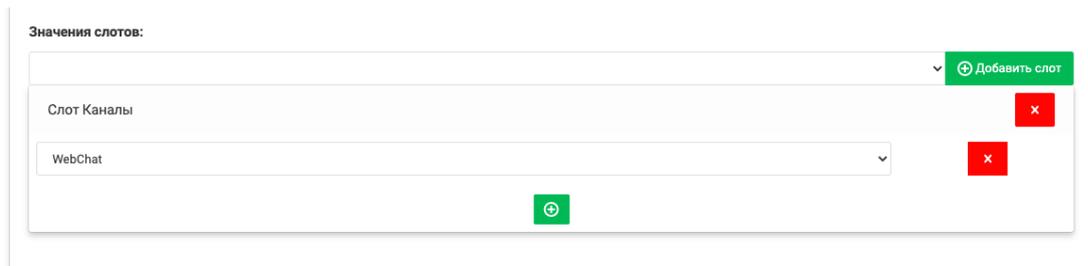


Рисунок 3.45

9. Для добавления ответа для другого канала, например, Viber, повторить действия, выбрав из выпадающего списка значение Viber (рис. 3.46 - 5.47).

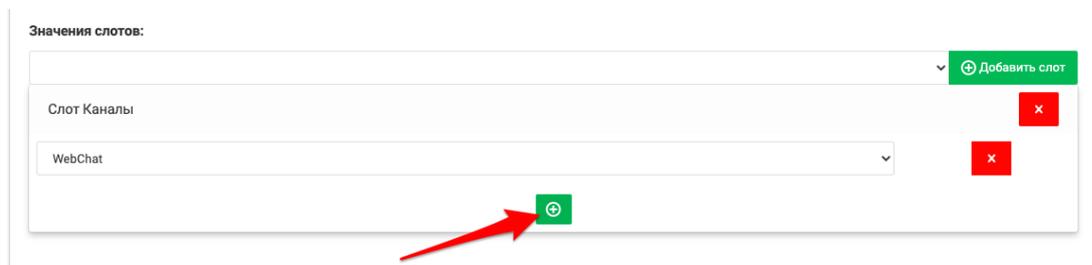


Рисунок 3.46

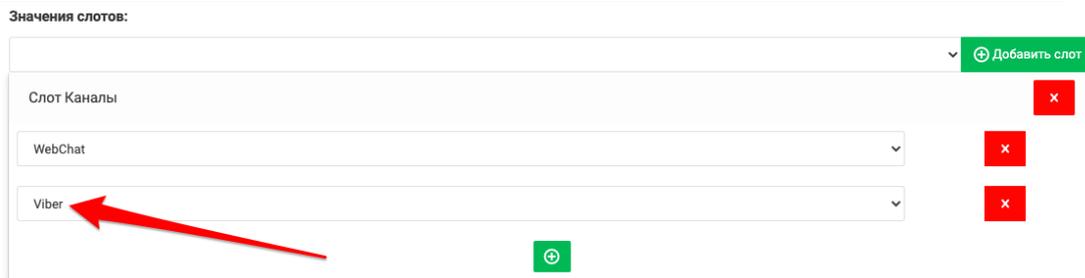


Рисунок 3.47

В результате выполненных действий в статье будет создано два ответа, под каждым из которых будет установлен слот. Под первым ответом, для канала вебчат, значение слота будет установлено как WebChat, а под вторым ответом, для канала вайбер, значение слота будет установлено как Viber.

10. Сохранить изменения в статье, выполнить активацию записи (см. п.5.9).

Когда система выберет в качестве ответа на вопрос Обратившегося статью, в которой есть разделение ответов на каналы, если вопрос Обратившегося пришел из веб-чата, ему будет отправлен первый ответ (под которым в качестве значения слота Каналы выбрано WebChat), а если вопрос Обратившегося придет из канала viber, то уйдет ответ, под которым установлен слот Каналы в значение Viber.

3.9 Активация новой записи в Базе знаний

1. После заполнения всех разделов или внесения изменений в них необходимо выполнить процедуру активации новой записи, для чего в любом разделе нажать «Создать черновик» (рис. 3.48). Появится надпись: «Запись успешно создана».

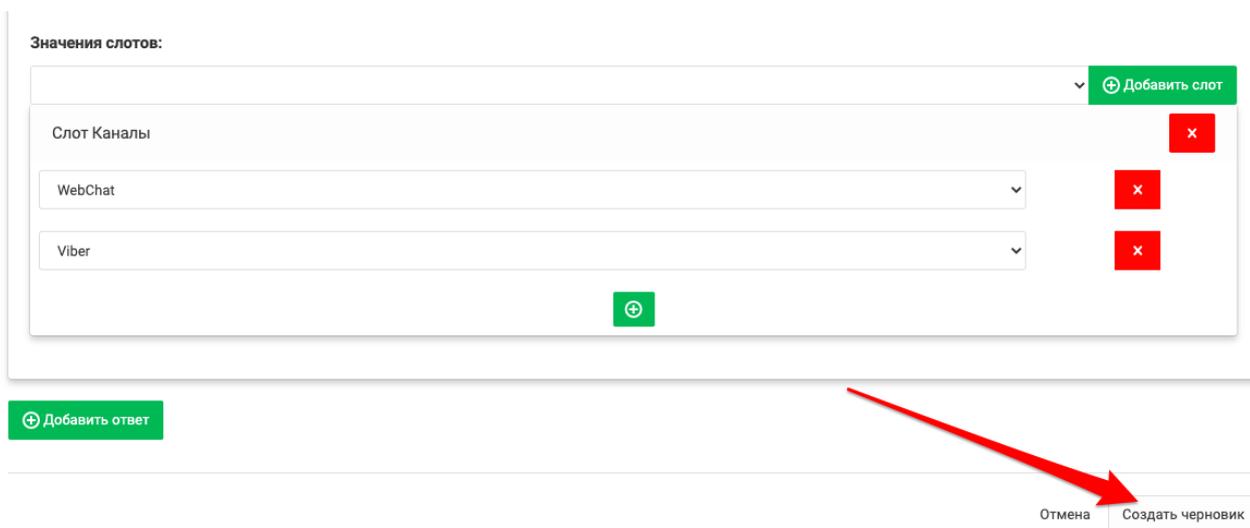


Рисунок 3.48

2. Далее необходимо вернуться в меню «Записи базы знаний», найти по имени или статусу (в таблице «Фильтр по статусам» выбрать «Черновик») новую запись и нажать на неё в «Название» (рис. 3.49). В появившемся редакторе записи нажать «Отправить на проверку» (при соответствующих правах этот пункт можно пропустить и сразу опубликовать статью) (рис. 3.50). Появится надпись - «Запись отправлена на проверку».

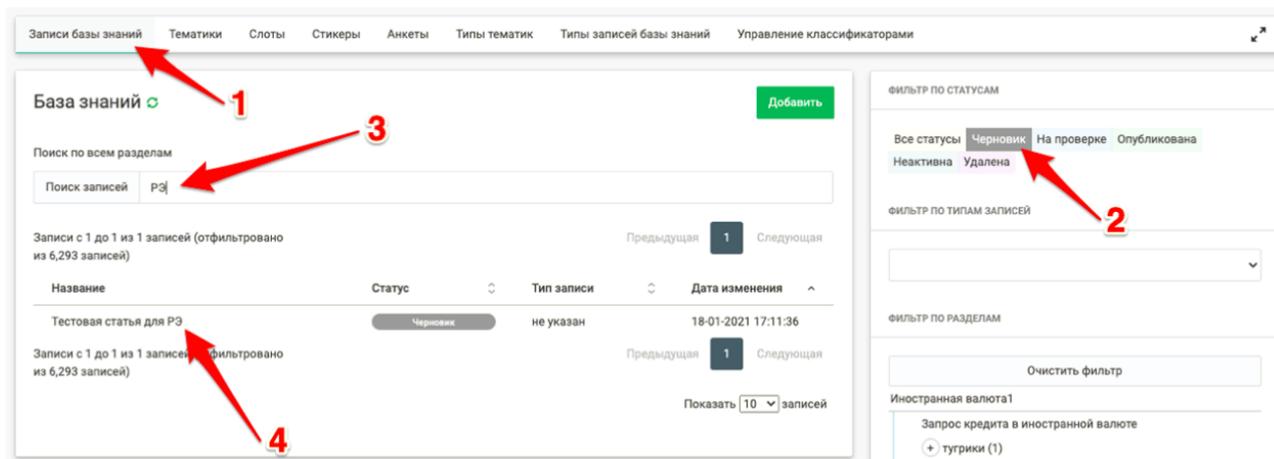


Рисунок 3.49

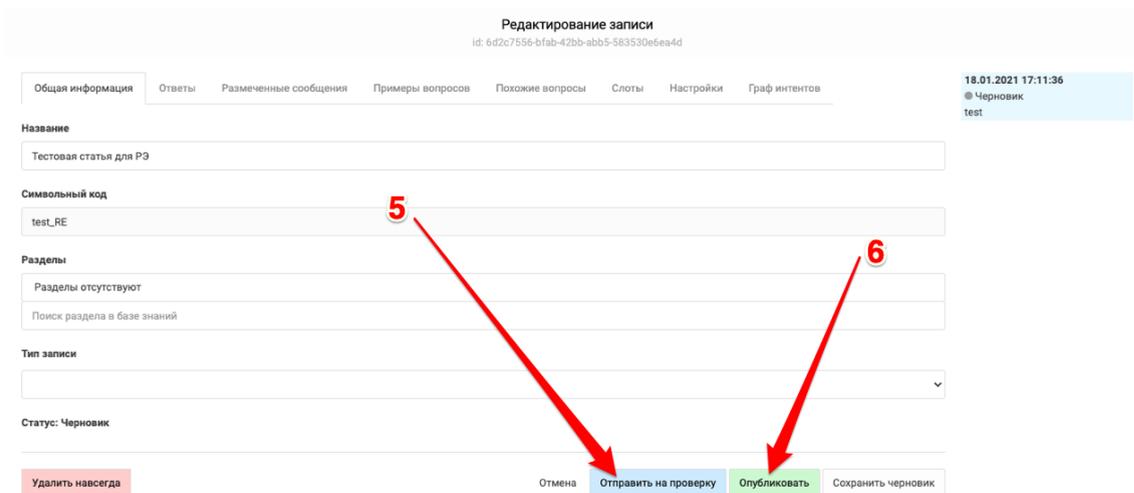


Рисунок 3.50

3. Далее по аналогии с п.2 находим новую запись со статусом «На проверке» и нажимаем на неё. В редакторе записи нажимаем «Опубликовать» (для этого должна быть установлена роль «Публикация статьи») (см. рис. 3.50). Появится надпись: «Запись успешно опубликована».

4. Далее можно внести изменения в запись в редакторе записи либо провести процедуру обучения Бота (описана в параграфе «Управление классификаторами»).

7. При наличии у пользователя роли "Удаление статьи из базы знаний навсегда" становится доступна кнопка "Удалить навсегда" (рис. 3.51), нажатие на которую приводит к физическому удалению статьи из Базы знаний, включая все ее версии. Восстановить статью, удаленную таким образом, нельзя.

Редактирование записи
id: 16fbd96d-3420-4314-ae4e-f48d0b1fa65f

Общая информация | Ответы | Размеченные сообщения | Примеры вопросов | Похожие вопросы | Слоты | Настройки

21.06.2021 20:37:54
● Черновик
nick2112@mail.ru

Название
тест_РЭ

Символьный код
test_re

Разделы
Разделы отсутствуют
Поиск раздела в базе знаний

Тип записи

Статус: Черновик

Удалить навсегда | Отмена | Отправить на проверку | Опубликовать | Сохранить черновик

Рисунок 3.51

В Системе КТ поддерживается версионность Базы знаний. Справа в таблице отображаются все состояния статьи – даты изменения и состояние. Статьи в этой таблице можно просмотреть и сравнить (внизу каждой версии надпись «Сравнить») с опубликованной (изменённая информация выделяется фоном), просмотреть предыдущие версии (рис. 3.52).

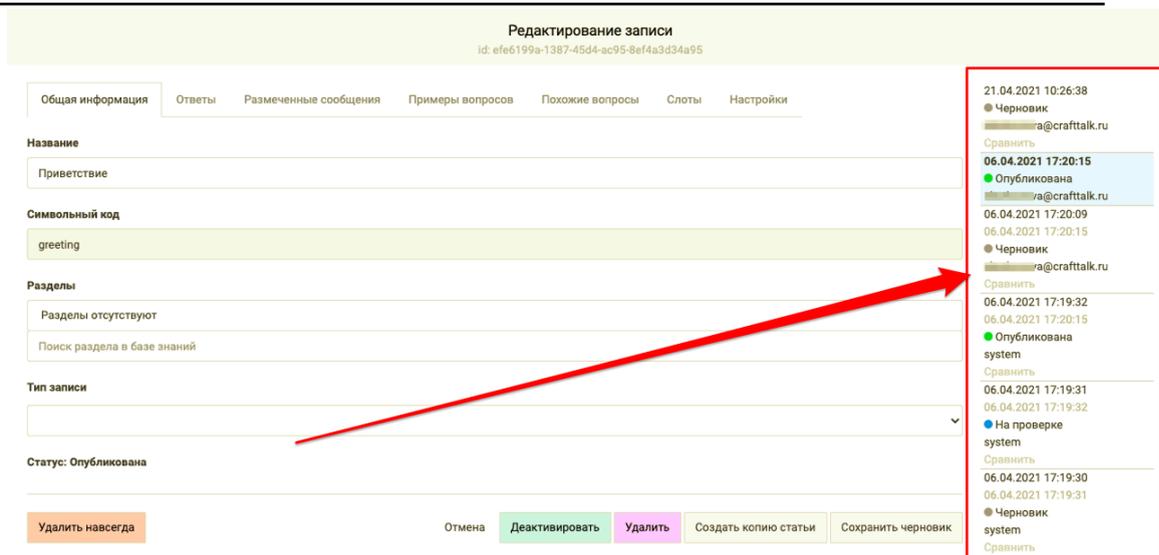


Рисунок 3.52

3.10 Редактирование записи

Нажатие в меню «Записи базы знаний» на строке таблицы с записью открывает вкладку «Редактирование записи», где доступен редактор БЗ, который позволяет искать, редактировать существующие статьи, а также удалять статьи Базы знаний. Под надписью «Редактирование записи» виден идентификатор, который создаётся автоматически при добавлении новой записи.

Справа в окне «Редактирование записи» видна вся история создания и изменений записи Базы знаний. История может быть просмотрена и статус записи изменён, а удалённая запись восстановлена.

Редактор содержит разделы для редактирования:

- Общая информация;
- Ответы;
- Размеченные сообщения;
- Примеры вопросов;
- Похожие вопросы;
- Слоты;
- Настройка.

3.11 Редактирование записи – «Общая информация»

В этом разделе возможно редактирование и изменение следующих параметров записи (рис. 3.53):

- название;

- символный код;
- разделы;
- тип записи (типичный вариант: ФЛ – физические лица, ЮЛ - юридические лица, Прочее).

Строка «Статус» - для информирования о текущем статусе записи.

Редактирование записи
id: efe6199a-1387-45d4-ac95-8ef4a3d34a95

Общая информация
Ответы
Размеченные сообщения
Примеры вопросов
Похожие вопросы
Слоты
Настройки

Название

Символьный код

Разделы

Разделы отсутствуют

Поиск раздела в базе знаний

Тип записи

Статус: Опубликована

Удалить навсегда
Отмена
Деактивировать
Удалить
Создать копию статьи
Сохранить черновик

21.04.2021 10:26:38

● Черновик

[Сравнить](#)

06.04.2021 17:20:15

● Опубликована

[Сравнить](#)

06.04.2021 17:20:09

06.04.2021 17:20:15

● Черновик

[Сравнить](#)

06.04.2021 17:19:32

06.04.2021 17:20:15

● Опубликована

[Сравнить](#)

06.04.2021 17:19:31

06.04.2021 17:19:32

● На проверке

[Сравнить](#)

06.04.2021 17:19:30

06.04.2021 17:19:31

● Черновик

[Сравнить](#)

Рисунок 3.53

3.12 Редактирование записи – «Ответы»

В этом разделе возможно редактирование и изменение следующих параметров записи (рис. 3.54):

- тип текста ответа (HTML ответы используются только для информации Операторов и канала вебчат; markdown - для мессенджеров и социальных сетей; простой текст - для любых каналов);
- изменение, добавление или удаление ответов;
- удаление и добавление кнопок;
- изменение значений слотов.

Общая информация | **Ответы** | Размеченные сообщения | Примеры вопросов | Похожие вопросы | Слоты | Настройки

Ответы

Ответ Здравствуйте! Чем я могу вам помочь?

Тип текста ответа

Простой текст HTML Markdown

Добавить шаблон

Здравствуйте! Чем я могу вам помочь?

Укорачивание ссылок:

Введите адрес ссылки

Вложения:

Введите адрес документа

Кнопки:

Значения слотов:

Рисунок 3.54

Для добавления текста ответа необходимо нажать на кнопку «Добавить ответ» (рис. 3.55). Нажатие этой кнопки приводит к появлению плашки со ссылкой «Ответ», которая разворачивает форму ввода нового ответа (состоит из двух страниц – ввода нового ответа и редактирования существующего) (рис. 3.56).

Общая информация | **Ответы** | Размеченные сообщения | Примеры вопросов | Похожие вопросы | Слоты | Настройки

Ответы

Ответ Здравствуйте! Чем я могу вам помочь?

Тип текста ответа

Простой текст HTML Markdown

Добавить шаблон

Здравствуйте! Чем я могу вам помочь?

Укорачивание ссылок:

Вложения:

Введите адрес документа

Кнопки:

Значения слотов:

Рисунок 3.55

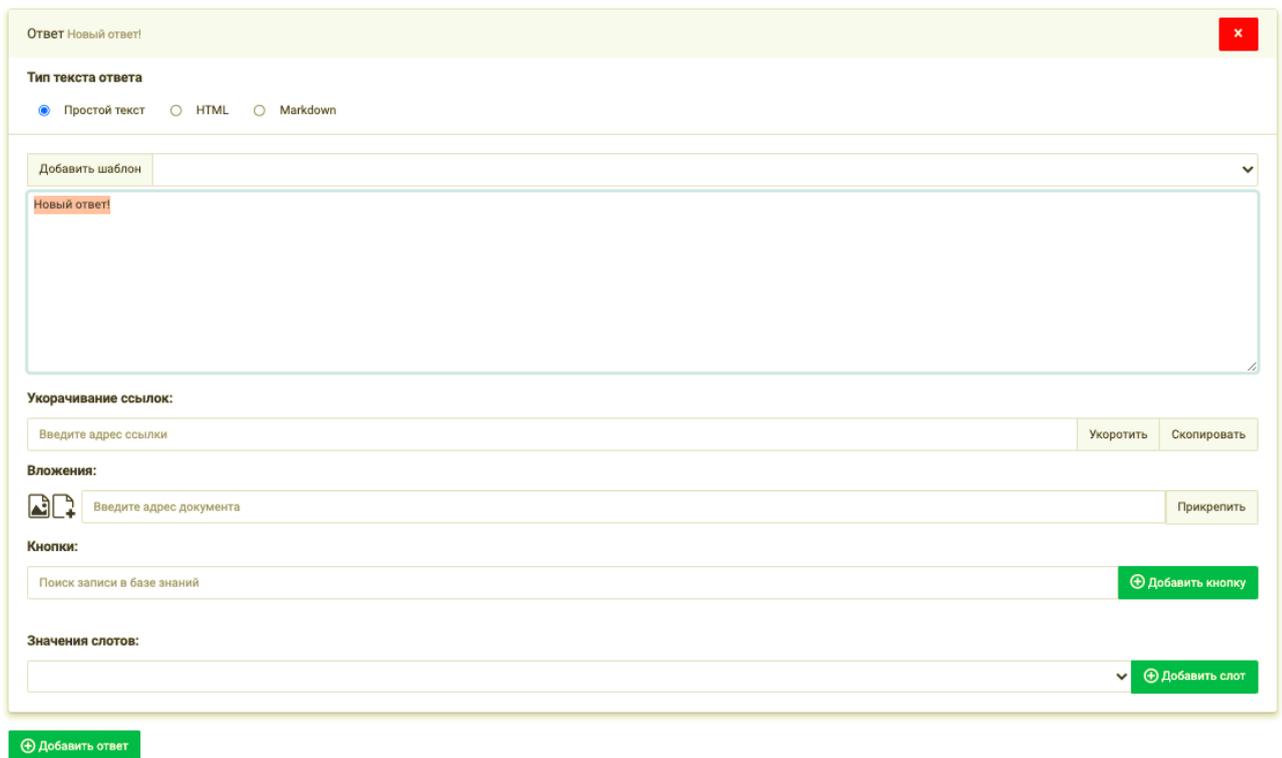


Рисунок 3.56

Ответ вводится в специальное поле «Ответы».

В разделе «Вложения» можно прикрепить картинку (нажать на значок-ссылку «Послать фото» и указать путь к нему в открывшемся диалоге либо просто перетащив выбранный файл в окно создания ответа), вложенный документ («Послать файл»), или вставить ссылку на документ на сайте и нажать кнопку «Прикрепить».

«Добавить кнопку» — это механизм добавления кнопки. «Кнопки» выводятся после текста ответа Обратившемуся и служат для быстрого перехода Обратившегося к другой информации, например, к другим статьям БЗ.

Для создания новой кнопки требуется ввести идентификатор существующей (или будущей) статьи в БЗ и название, которое будет отображаться на самой кнопке («Текст кнопки»), кроме того, есть возможность добавить синонимы к кнопке (см. п. 3.12.1).

Для удобства поиска записей в БЗ, на которые будут ссылаться кнопки, предусмотрена строка поиска (рис. 3.57). В нее достаточно ввести название записи. В случае успешного поиска будет выведена подсказка, на которую достаточно кликнуть мышкой, и поля идентификатор и текст кнопки заполнятся автоматически.

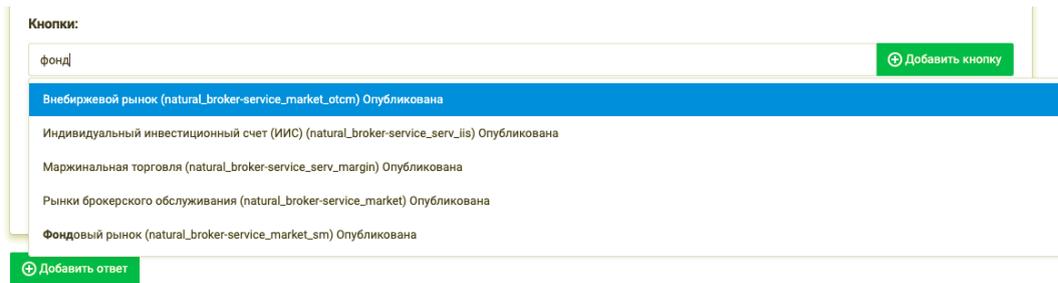


Рисунок 3.57

В случае необходимости кнопку можно удалить из ответа. Для удаления кнопки требуется нажать на красный крестик справа от строки с кнопкой и подтвердить удаление в открывшемся диалогом окне.

Следует помнить, что количество ответов к одной статье базы знаний неограниченно. Механизм добавления новых ответов аналогичен описанному выше.

В случае необходимости ответ можно удалить из записи. Для удаления ответа надо нажать на кнопку удаления  справа от ответа и подтвердить удаление ответа в открывшемся диалоговом окне.

3.12.1 Редактирование записи – «Ответы». Синонимы для кнопки

Добавление синонимов к кнопке возможно в случае, если выбран тип слота "Словарь".

Синонимы берутся из слотов. Для каждого предложения и каждой группы синонимов необходимо проверить, есть ли вхождение одного из синонимов в предложение. Если есть, то необходимо добавить в обучающую выборку примеры, где в предложение найденное вхождение заменено каждым синонимом.

Инструкция по добавлению синонимов:

1. Необходимо в Базе знаний в разделе Слоты создать Слот типа "Словарь" (рис. 3.58 – 5.59). Указываем любой идентификатор и запоминаем его (например «slot_sinonim_1»).

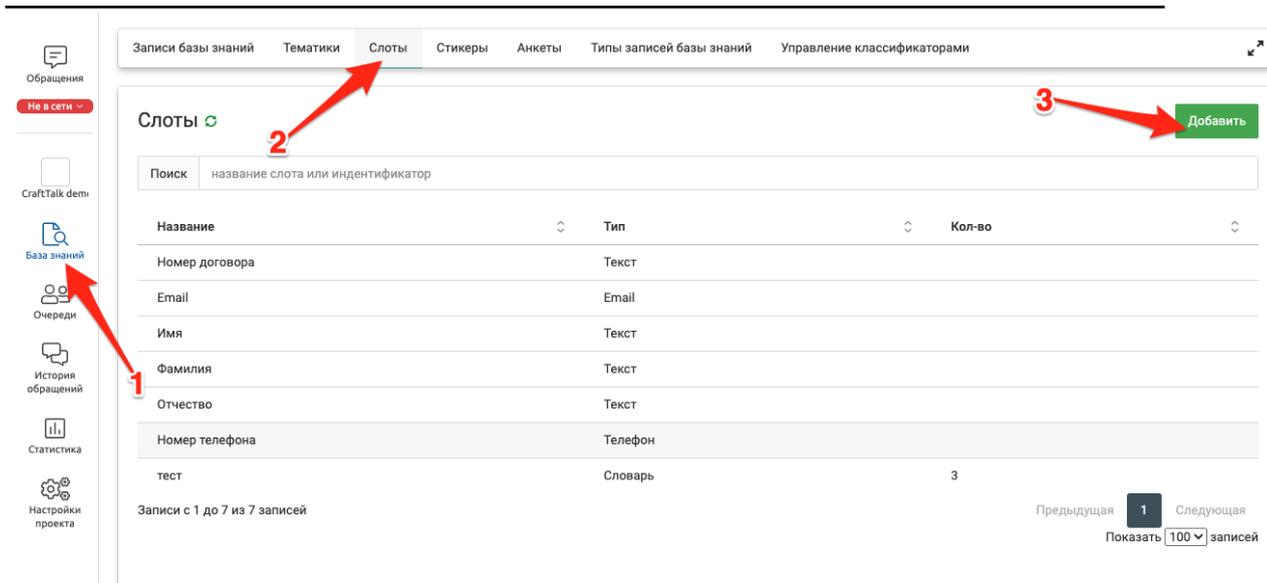


Рисунок 3.58

Название

Идентификатор

Тип слота

Предзаполнять слот

Пропускать классификацию при уточнении

Показывать кнопку отмены

Не показывать кнопки словаря

Значения словаря

Вопрос

Второй вопрос

Вопрос для получения следующего значения

Рисунок 3.59

2. В качестве значений Словаря используется Ключевое слово (рис. 3.60 1) и в графу “Синонимы” вносим слова-синонимы в данному слову (рис. 3.60 2).

Значения словаря

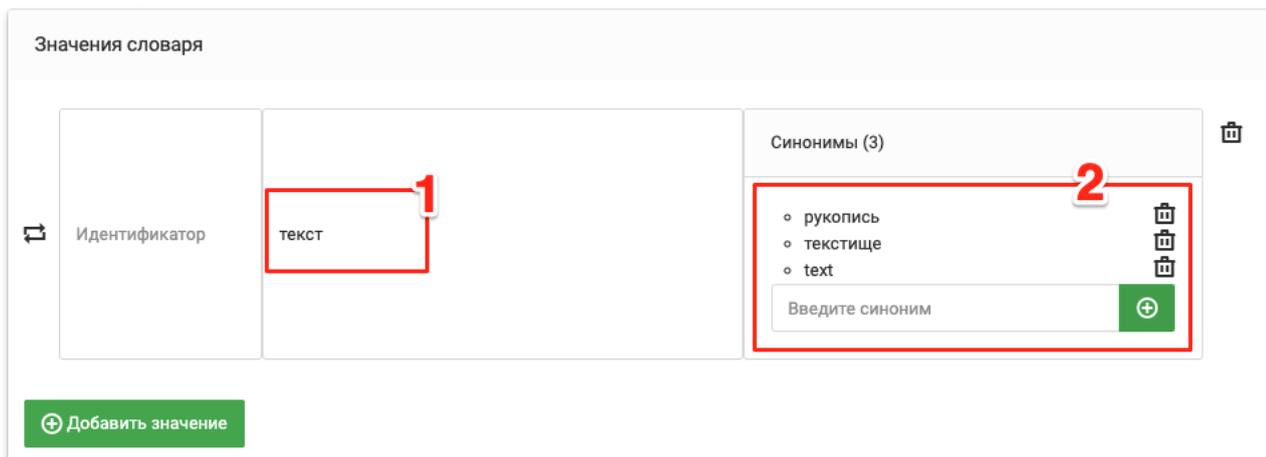


Рисунок 3.60

3. Сохраняем слот.

4. Переходим к настройке классификаторов (см. [классификаторы](#)). База знаний, раздел - Управление классификаторами (рис. 3.61).

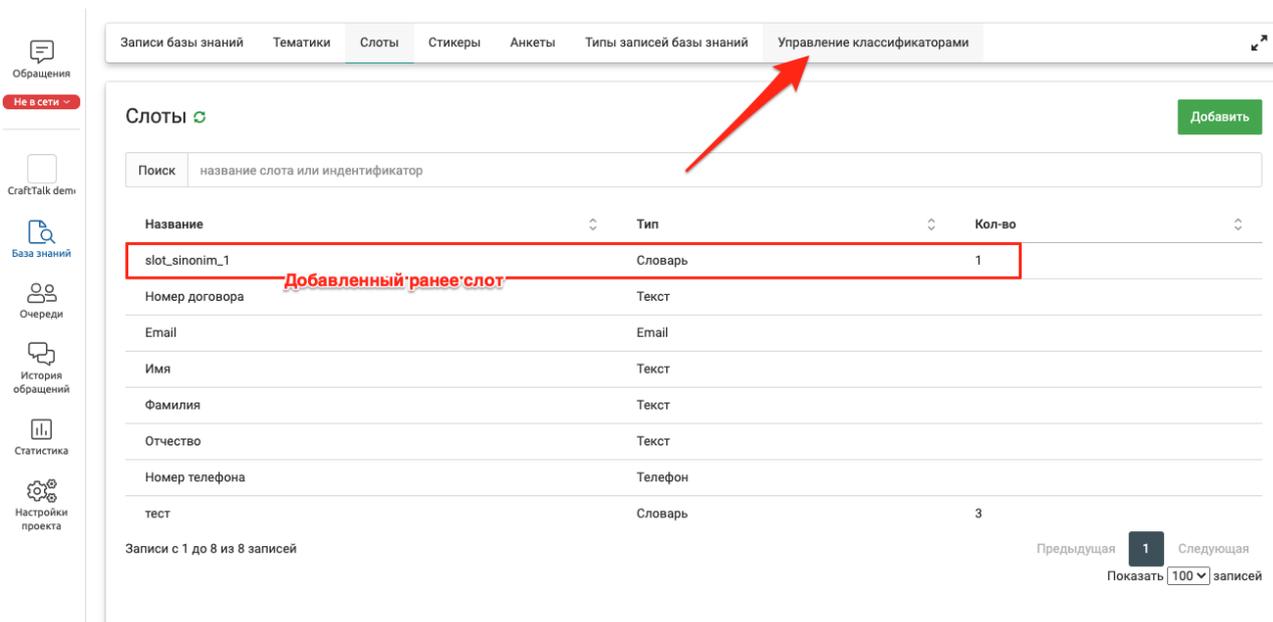


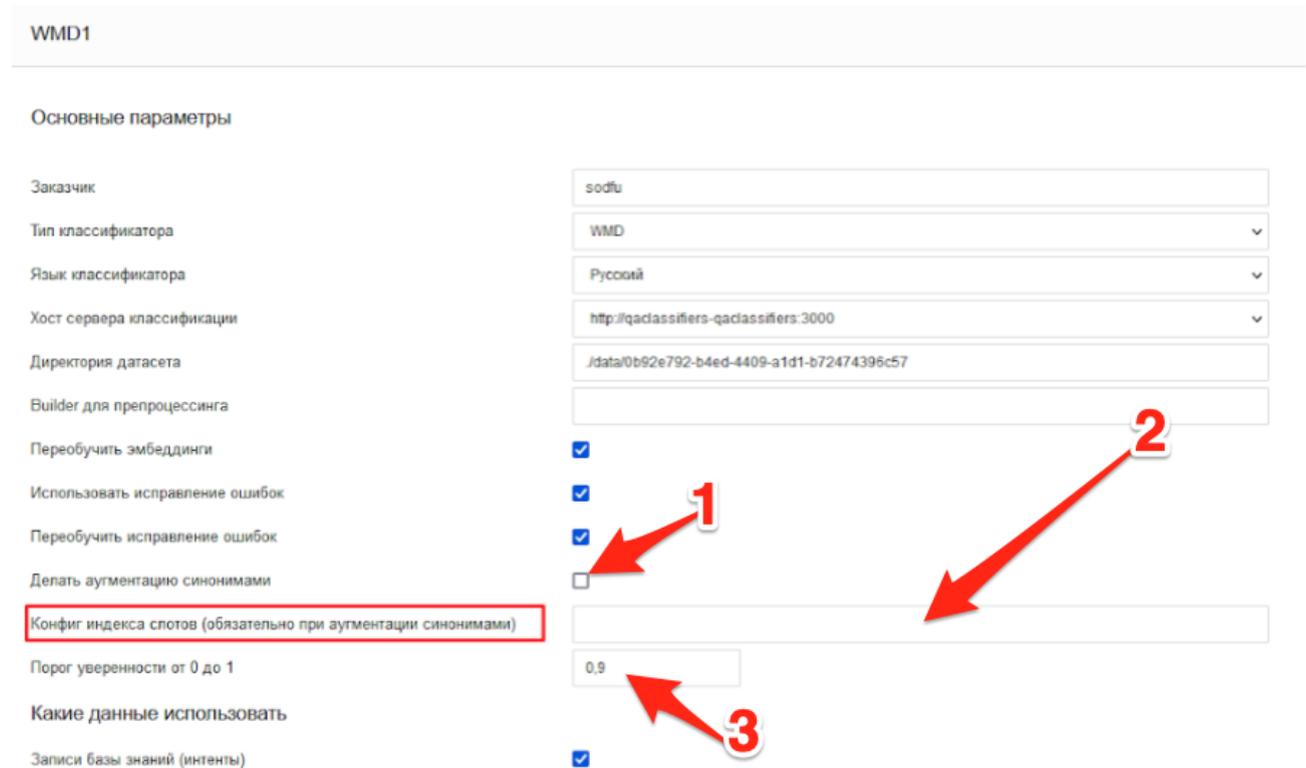
Рисунок 3.61

5. В настройках классификатора необходимо установить галочку в поле “Делать аугментацию синонимами” (рис. 3.62 1).

6. В поле “Конфиг индекса слотов (обязательно при аугментации синонимами)” необходимо указать строку следующего вида: http://<host>:<port>/<index_name>/<type> (рис. 3.62 2).

Пример - http://elasticsearch-elasticsearch-es-http.elastic:9200/slot_sis/slot.

7. Далее необходимо указать “Порог уверенности от 0 до 1” (рис. 3.62 3) - поиск слова в предложении нечеткий, как и у слотов. Граница уверенности регулируется через slot_threshold: 0 - заменять все, 1 - заменять только точное совпадение. Значение по умолчанию в системе - 0.9.



WMD1

Основные параметры

| | |
|--|--|
| Заказчик | sodfu |
| Тип классификатора | WMD |
| Язык классификатора | Русский |
| Хост сервера классификации | http://qaclassifiers-qaclassifiers:3000 |
| Директория датасета | /data/0b62e792-b4ed-4409-a1d1-b72474396c57 |
| Builder для препроцессинга | |
| Переобучить эмбединги | <input checked="" type="checkbox"/> |
| Использовать исправление ошибок | <input checked="" type="checkbox"/> |
| Переобучить исправление ошибок | <input checked="" type="checkbox"/> |
| Делать аугментацию синонимами | <input type="checkbox"/> |
| Конфиг индекса слотов (обязательно при аугментации синонимами) | |
| Порог уверенности от 0 до 1 | 0,9 |
| Какие данные использовать | |
| Записи базы знаний (интенты) | <input checked="" type="checkbox"/> |

Рисунок 3.62

3.13 Редактирование записи – «Размеченные сообщения»

Помимо явно указанных формулировок, интеллектуальный помощник и Бот используют для обучения информацию, накопленную в процессе работы Операторов с Базой знаний (выбора Оператором статей для отправки в качестве ответов на вопрос) и в процессе ручной разметки при работе с историей обращений, которая хранится в разделе «Размеченные сообщения» (рис. 3.63).

Здесь перечислены вопросы, которые Администратор Базы знаний или Оператор сочли подходящими для того, чтобы отправить в канал клиента информацию, хранящуюся в ответе данной статьи.

На вкладке «Размеченные сообщения» можно удалить вопросы Обратившегося, на которые может быть дан ответ из текущей записи БЗ. Эти вопросы попадают в данный список после разметки диалога Бизнес Администратором в меню «Поиск обращений». Для удаления некоторых вопросов необходимо разметить галочками удаляемые вопросы слева от их текста и нажать

на кнопку «Удалить отмеченные» вверху справа от строки поиска по вопросам. При нажатии на кнопку «Удалить все» будут удалены все вопросы на текущей вкладке «Размеченные сообщения». Если редактировалось содержимое только вкладки "Размеченные сообщения" (и/или "Похожие вопросы"), то сохранять изменения в статье не требуется. Достаточно нажать на кнопку "Отменить".

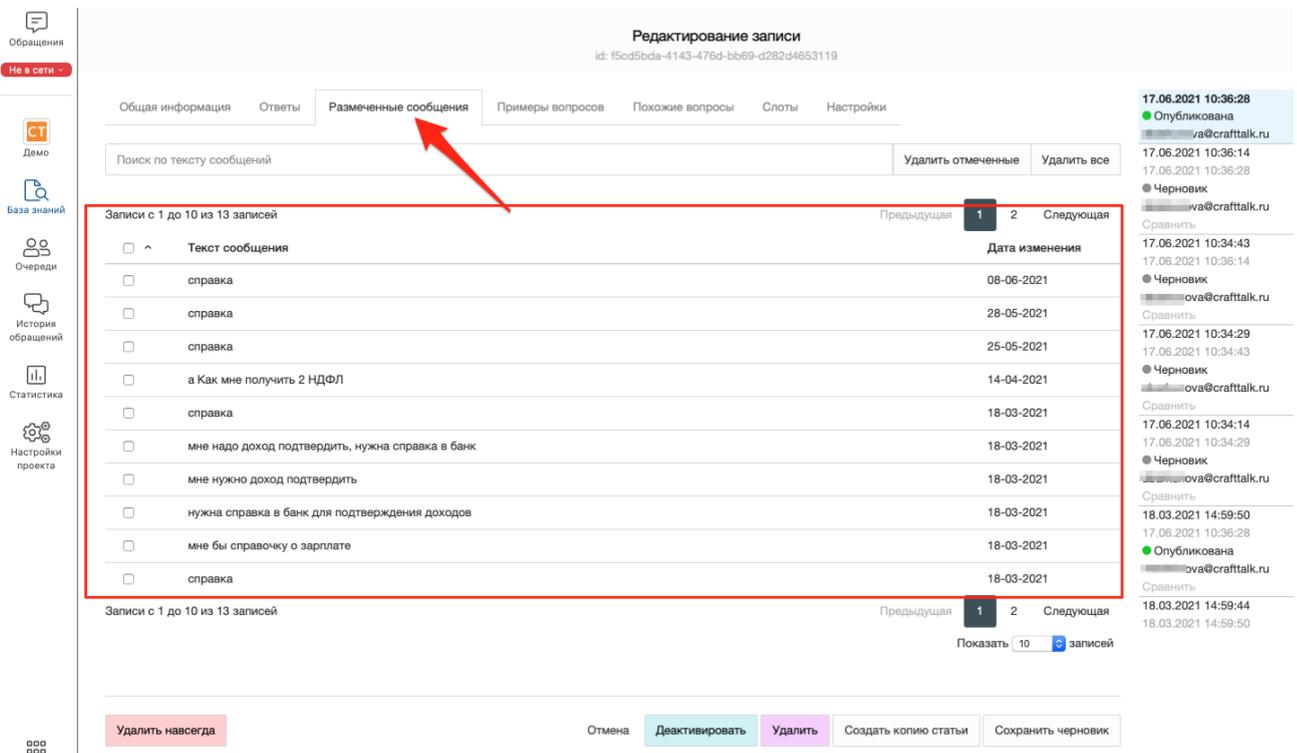


Рисунок 3.63

Запись можно деактивировать, не удаляя, нажав «Деактивировать», а также создать копию нажав «Создать копию статьи».

3.14 Редактирование записи – «Примеры вопросов»

В этом разделе можно отредактировать уже введенные примеры вопросов, отредактировав их текст, или добавить примеры вопросов. Также можно внести сразу несколько вопросов в «Список вопросов» (рис. 3.64).

Для редактирования текста вопроса достаточно мышкой выделить вопрос и внести необходимые коррективы.

Добавление вопросов завершается нажатием кнопки . Для удаления вопроса – нажать на кнопку  справа от вопроса.

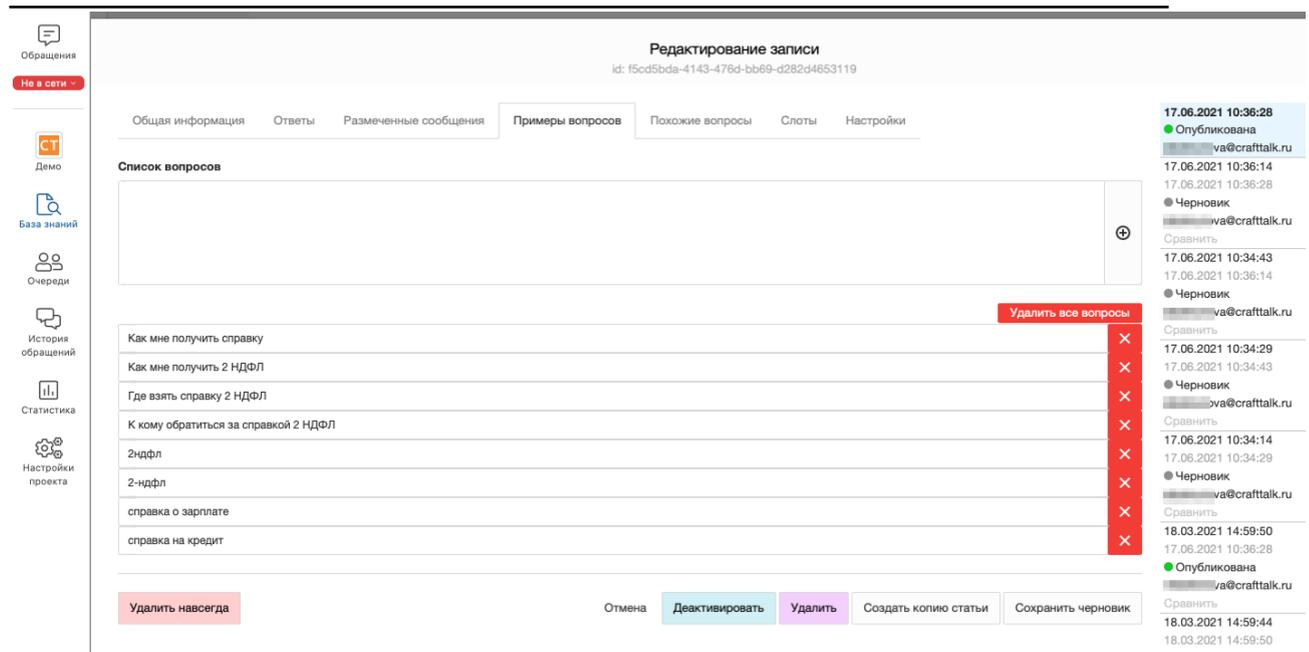


Рисунок 3.64

Для лучшего обучения требуется введение как можно большего числа вопросов в различных формулировках, при этом не требуется введения вопросов в различных падежах и числе. Количество вопросов не ограничено.

3.15 Редактирование записи – «Похожие вопросы»

Система КТ анализирует все ведущиеся диалоги и может предложить в качестве похожих вопросов подходящие с точки зрения искусственного интеллекта вопросы клиента, которые были заданы в диалогах для данной статьи. Они находятся на вкладке “Похожие вопросы” (рис. 3.65).

Администратор может выбрать подходящие с его точки зрения вопросы из всех предложенных. Отмеченные вопросы можно затем увидеть во вкладке “размеченные сообщения” см. рис..63, и в дальнейшем они будут использоваться для обучения бота.

Похожие вопросы используются Ботом и интеллектуальным помощником для поиска ответов на вопросы, не соответствующие списку в «Примерах вопросов». Постепенное расширение (обучение) этого раздела ускоряет процесс получения ответа на вопросы Обратившегося и увеличивает процент вероятности правильности ответа.

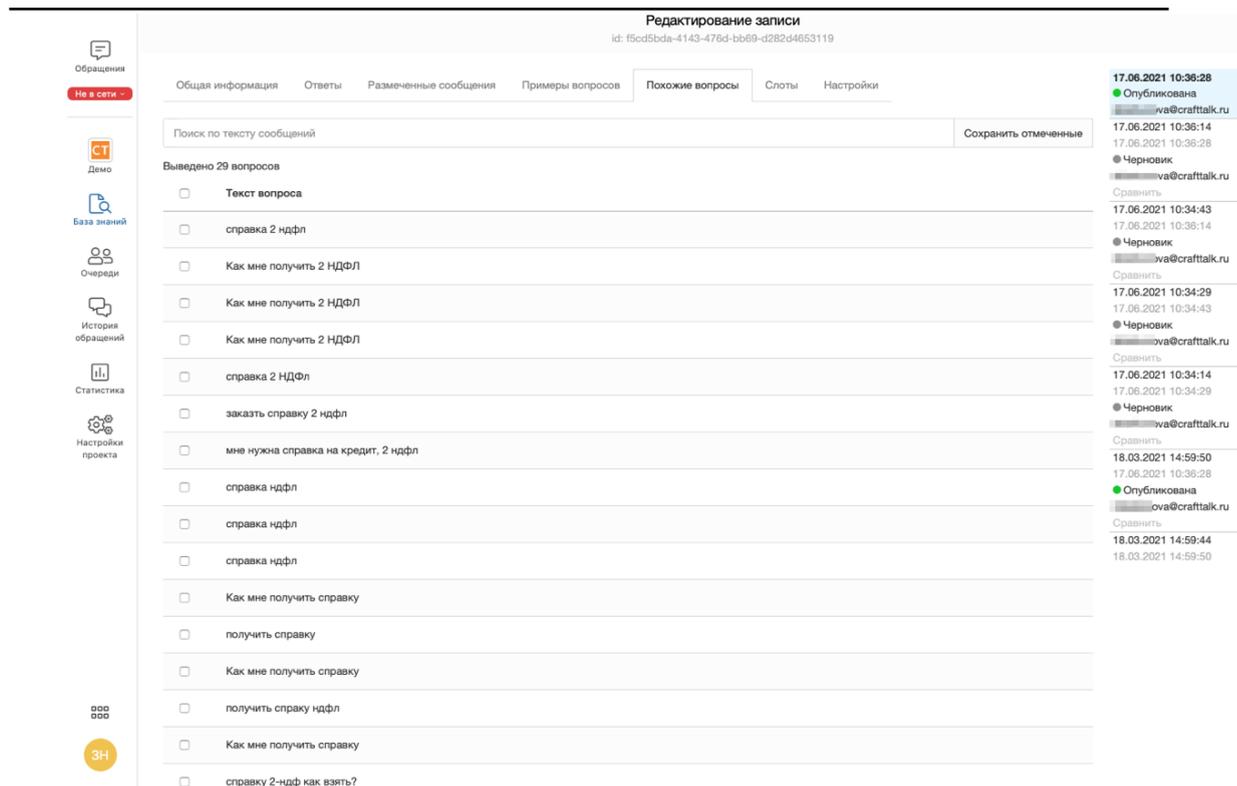


Рисунок 3.65

Меню «Похожие вопросы» заполняется поиском похожих вопросов по тексту сообщений. В строке поиска вводится текст похожего вопроса. При этом в таблице «Текст вопроса» отсортировываются похожие вопросы для выбора. После разметки галочками похожих вопросов нажать «Сохранить отмеченные» (рис. 3.66).

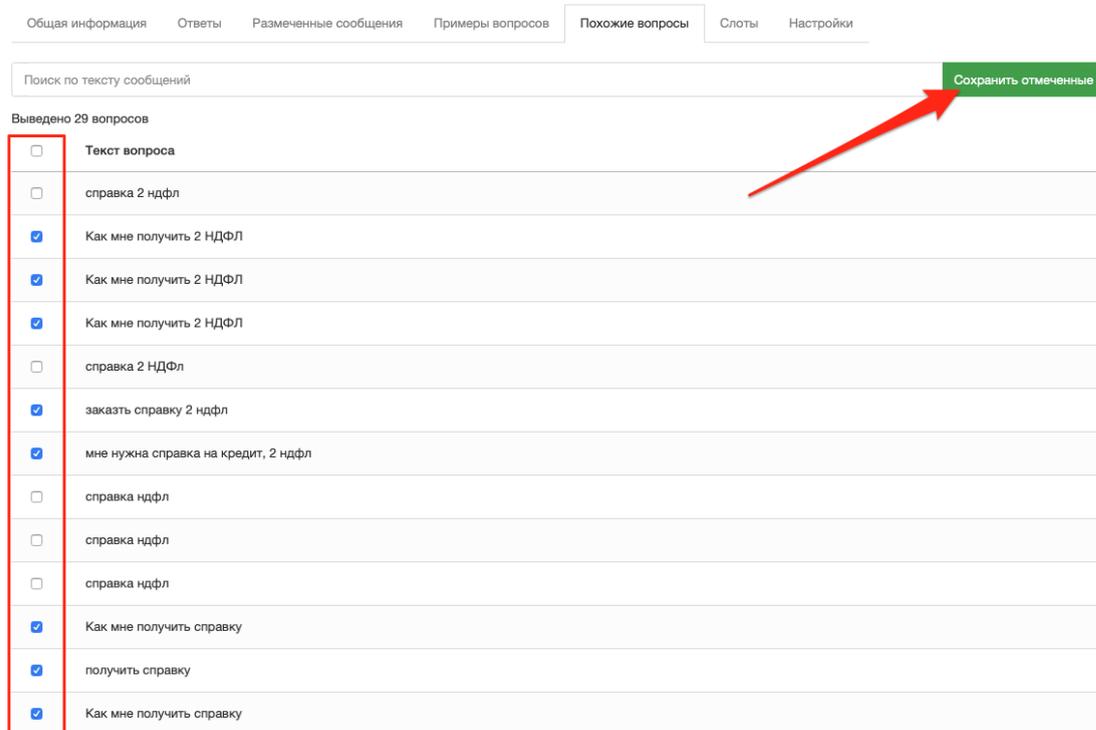


Рисунок 3.66

Должна появиться надпись: «Разметка сохранена». Размеченные вопросы попадут в таблицу на вкладке «Размеченные сообщения» и будут использованы для обучения классификатора. Если редактировалось содержимое только вкладки «Похожие вопросы» (и/или «Размеченные сообщения»), то сохранять изменения в статье не требуется. Достаточно нажать на кнопку «Отменить».

3.16 Редактирование записи – «Слоты»

В данном разделе возможно дополнительно определить привязку записи к слотам либо удалить существующие привязки (рис. 3.67).

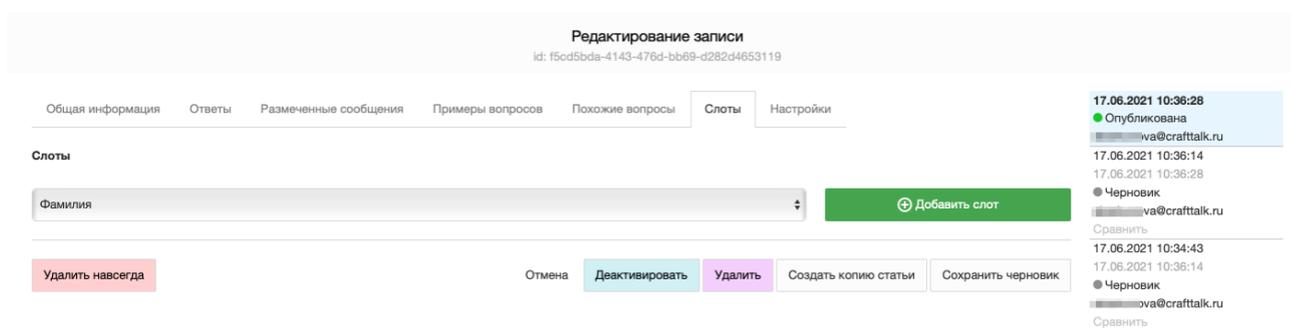


Рисунок 3.67

3.17 Активация отредактированной записи

После внесения изменений необходимо выполнить процедуру активации записи, для чего в любом разделе нажать «Создать черновик» (рис. 3.68). Появится надпись: «Запись успешно создана» (рис. 3.69).

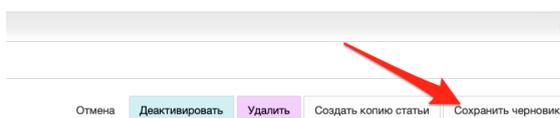
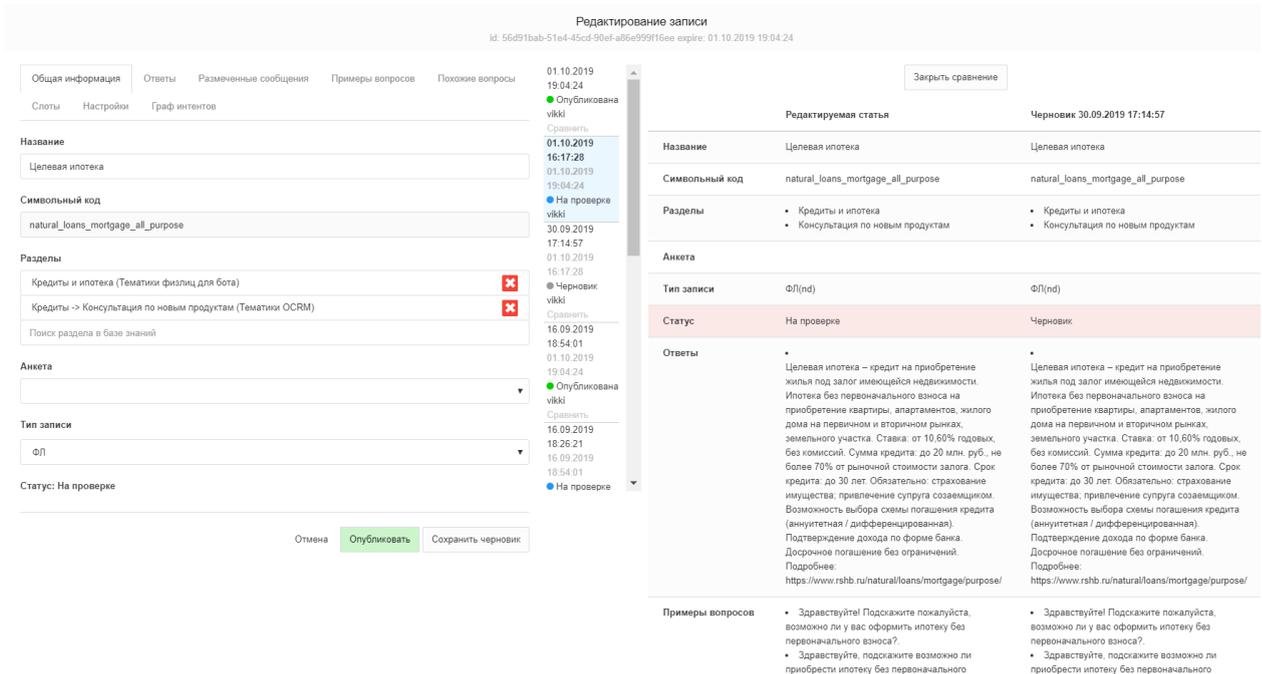


Рисунок 3.68

Далее необходимо вернуться в меню «Записи базы знаний», найти по имени или статусу (в таблице «Фильтр по статусам» выбрать «Черновик») новую запись и нажать на неё в «Название». В появившемся редакторе записи нажать «Отправить на проверку». Появится надпись: «Запись отправлена на проверку». Далее точно также находим новую запись со статусом «На проверке» и нажимаем. В редакторе записи нажимаем «Опубликовать». Появится надпись: «Запись успешно опубликована». Далее необходимо провести процедуру обучения Бота (описана в

параграфе «Управление классификаторами»). При определенных правах запись в Базе знаний можно опубликовать минуя этап проверки.



Редактирование записи
id: 56d91bab-51e4-45cd-90af-a86e999f16ee expire: 01.10.2019 19:04:24

01.10.2019 19:04:24
● Опубликована вики

Сравнить
01.10.2019 16:17:28
01.10.2019 19:04:24
● На проверке вики
30.09.2019 17:14:57
01.10.2019 16:17:28
● Черновик вики

Сравнить
16.09.2019 18:54:01
01.10.2019 19:04:24
● Опубликована вики

Сравнить
16.09.2019 18:26:21
16.09.2019 18:54:01
● На проверке

Закреть сравнение

| Редактируемая статья | Черновик: 30.09.2019 17:14:57 |
|----------------------|---|
| Название | Целевая ипотека |
| Символьный код | natural_loans_mortgage_all_purpose |
| Разделы | <ul style="list-style-type: none"> Кредиты и ипотека Консультация по новым продуктам |
| Анкета | <ul style="list-style-type: none"> Кредиты и ипотека Консультация по новым продуктам |
| Тип записи | ФЛ(нд) |
| Статус | На проверке |
| Статус | Черновик |
| Ответы | <ul style="list-style-type: none"> Целевая ипотека – кредит на приобретение жилья под залог имеющейся недвижимости. Ипотека без первоначального взноса на приобретение квартиры, апартаментов, жилого дома на первичном и вторичном рынках, земельного участка. Ставка: от 10.60% годовых, без комиссий. Сумма кредита: до 20 млн. руб., не более 70% от рыночной стоимости залога. Срок кредита: до 30 лет. Обязательно: страхование имущества, привлечение супруга созаемщиком. Возможность выбора схемы погашения кредита (аннуитетная / дифференцированная). |
| Примеры вопросов | <ul style="list-style-type: none"> Здравствуй! Подскажите пожалуйста, возможно ли у вас оформить ипотеку без первоначального взноса? Здравствуйте, подскажите возможно ли приобрести ипотеку без первоначального |

Отмена Опубликовать Сохранить черновик

Рисунок 3.69

Кроме этого, запись можно «удалить» и «деактивировать» т.е. исключить из выдачи и процесса обучения классификатора.

Если в интент (статью БЗ) вносить изменения только в разметку (это вкладки "размеченные сообщения" и "похожие вопросы"), то можно не сохранять изменения по стандартному сценарию (черновик - на проверке - опубликовано), а просто закрыть интент (нажать "отмена") и изменения сохранятся.

В Системе КТ поддерживается версияность Базы знаний. Справа в таблице отображаются все состояния статьи – даты изменения и состояние. Статьи в этой таблице можно просмотреть и сравнить с опубликованной (изменённая информация выделяется фоном), просмотреть предыдущие версии.

3.18 Вкладка «Тематики»

Меню «Тематики» (отображаются на фильтрах с правой стороны экрана) позволяет соотнести создаваемую статью БЗ с существующей тематикой (или создать новую тематику при необходимости) (рис. 3.70).

Для создания новой тематики требуется открыть соответствующий раздел или подраздел, кликнув на ссылку “Тематики”, нажать на кнопку «Создать тематику» и заполнить поля открывшейся формы создания тематик (рис. 3.71).

В поле «Настройка параметров доступа» выбрать (нажав «треугольник» в строке) «Любая линия».

В случае успешного создания новой Тематики Базы знаний (кнопка «Создать») в левом нижнем углу всплывет уведомление. В противном случае будет выведено уведомление о невозможности создания новой тематики в Базе знаний.

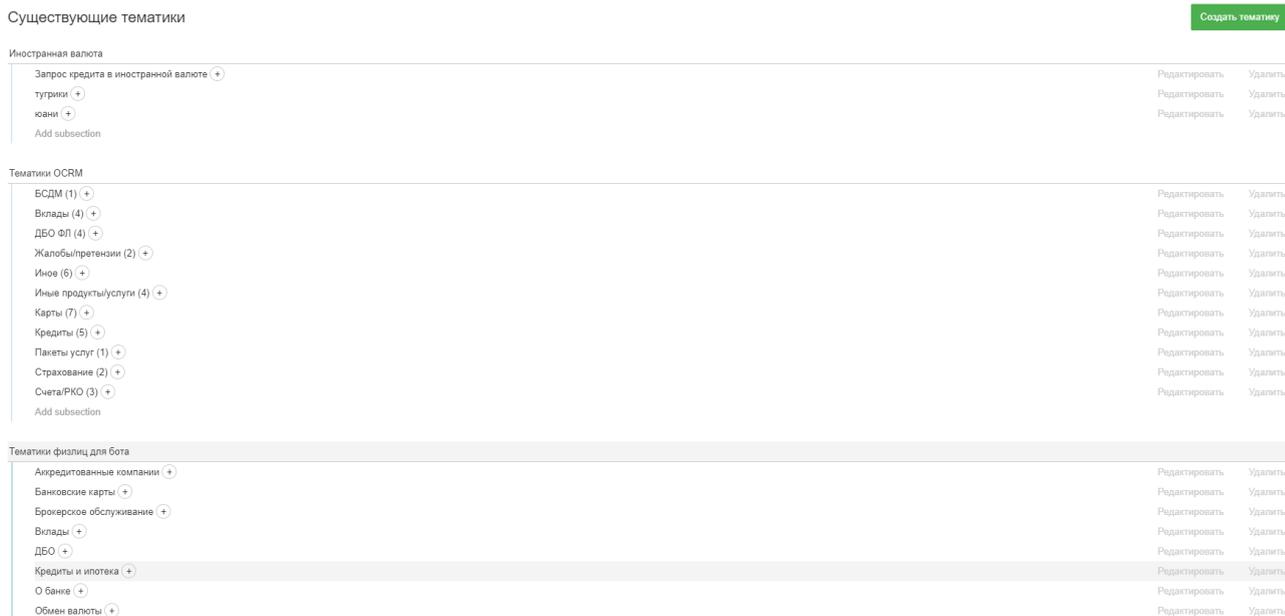


Рисунок 3.70

Создание тематики

Название

Символьный код

Настройки параметров доступа

Линия

▼

Импорт разделов тематики

Выберите файл для импорта разделов

Отмена

Рисунок 3.71

После сохранения Тематика появится в таблице записей и отобразится на экране (см. рис. 3.69). В случае отсутствия записи в таблице можно нажать на кнопку обновления  в правом верхнем углу формы или F5.

Далее в тематику должны быть внесены подразделы (тематики) созданной основной тематики.

После заполнения строк редактора разделов и нажатия «Создать» должна появиться надпись о создании раздела либо невозможности его создания.

Вновь внесенная Тематика сразу же становится доступна для полнотекстового поиска в справочнике на панели справа в «Записи базы знаний».

В строку поиска вводится слово для поиска. Результат поиска может быть отфильтрован по типу, доступному в выпадающем списке.

3.19 Вкладка «Типы записей базы знаний»

Типы записей Базы знаний можно создавать, редактировать и удалять для этого перейти во вкладку «Типы записей базы знаний» (рис. 3.72).

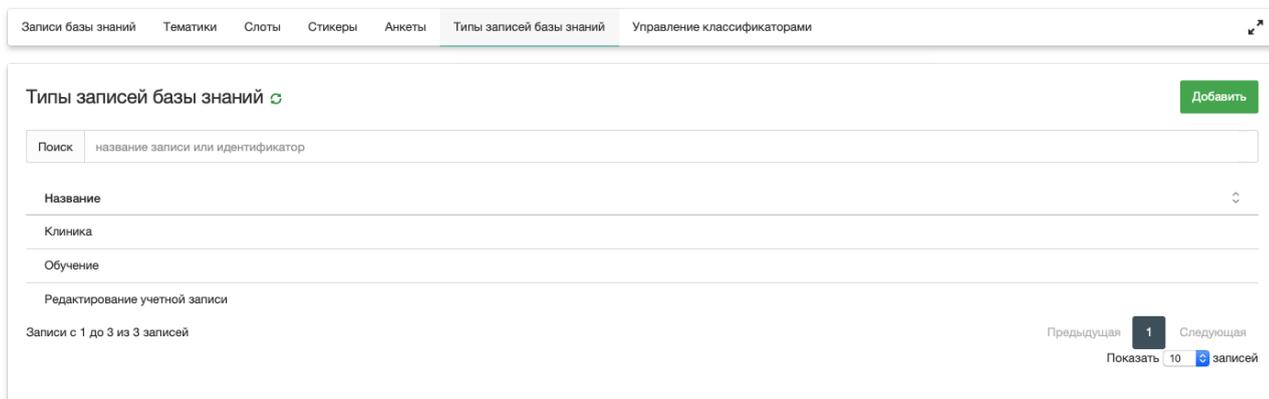


Рисунок 3.72

Поле «Тип записи базы знаний» проставляется для разделения статей БЗ. Например, с помощью данного признака в таблице с записями можно отсортировать интенты с ответами на вопросы пользователей, со стандартными вспомогательными ответами (например, приветствие/прощание/вопросы о погоде и т.п.).

Для добавления нового типа записи нажать «Добавить». В появившейся вкладке заполнить строки и нажать «Создать» (рис. 3.73).



Рисунок 3.73

Редактирование и удаление типа записи идентично аналогичным действиям с записями базы знаний.

Для редактирования или удаления - нажать в таблице на названии записи и, в появившейся вкладке, выполнить требуемые операции, после чего нажать «Сохранить» (рис. 3.74).

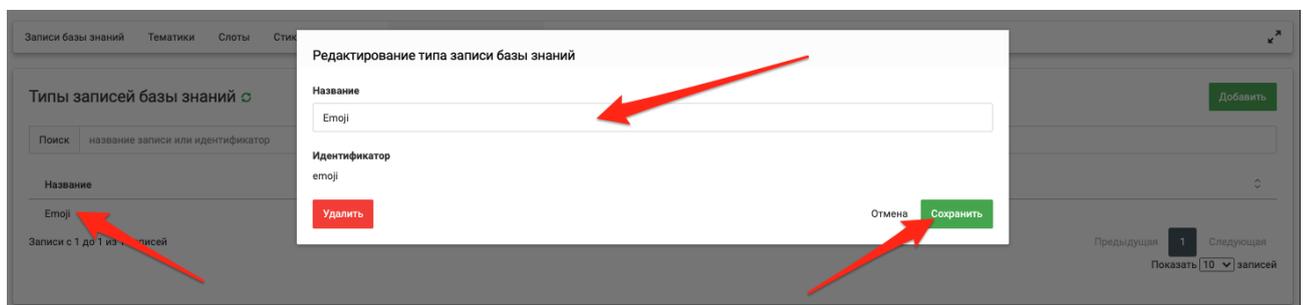


Рисунок 3.74

Редактирование и удаление типа записи идентично аналогичным действиям с записями Базы знаний.

Для создания типа записи требуется нажать на кнопку «Добавить» и заполнить два поля открывшейся формы — название нового типа и его идентификатор.

3.20 Вкладка «Слоты»

Во вкладке для просмотра и поиска возможна сортировка существующих слотов по названию, типу, количеству (значений в слоте) (рис. 3.75).

Платформа поддерживает концепцию сущностей, которые используются для заполнения слотов в процессе диалога. Слоты могут храниться в контексте диалога, пока он не завершен. Сущность – это единица информации, такая как имя, фамилия, дата, адрес электронной почты и так далее.

Поддерживаются следующие типы сущностей:

- целое число;
- текст;
- телефон;
- с проверкой регулярным выражением (с возможностью автоматического извлечения из запроса пользователя);
- словари (с возможностью рендеринга значений в виде кнопок и распознавания текстового ввода и автоматического извлечения значения из запроса пользователя с нечетким совпадением);
- местоположение.

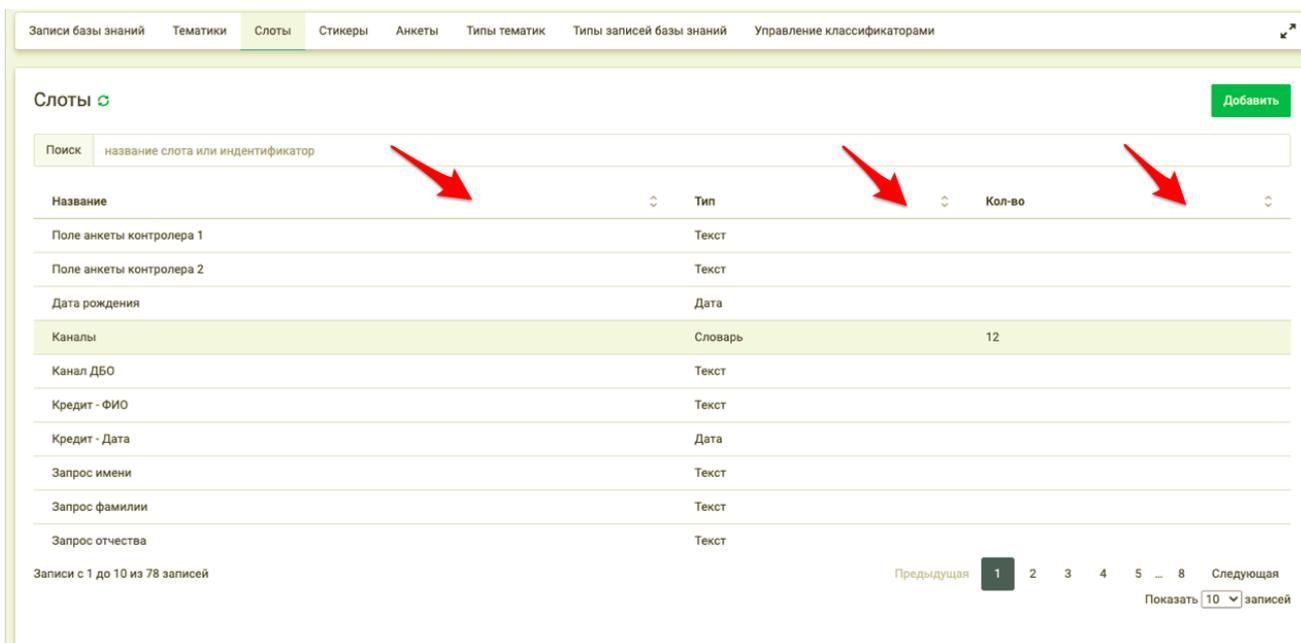


Рисунок 3.75

Помимо явно конфигурируемых слотов, платформа автоматически подставляет значения некоторых системных слотов – например идентификатор канала, что позволяет использовать эту информацию для ветвления диалога, специализации ответов и для внешних агентов.

Во вкладке «Слоты» есть возможность создания новых слотов, для этого необходимо нажать кнопку «Добавить» (рис. 3.76).

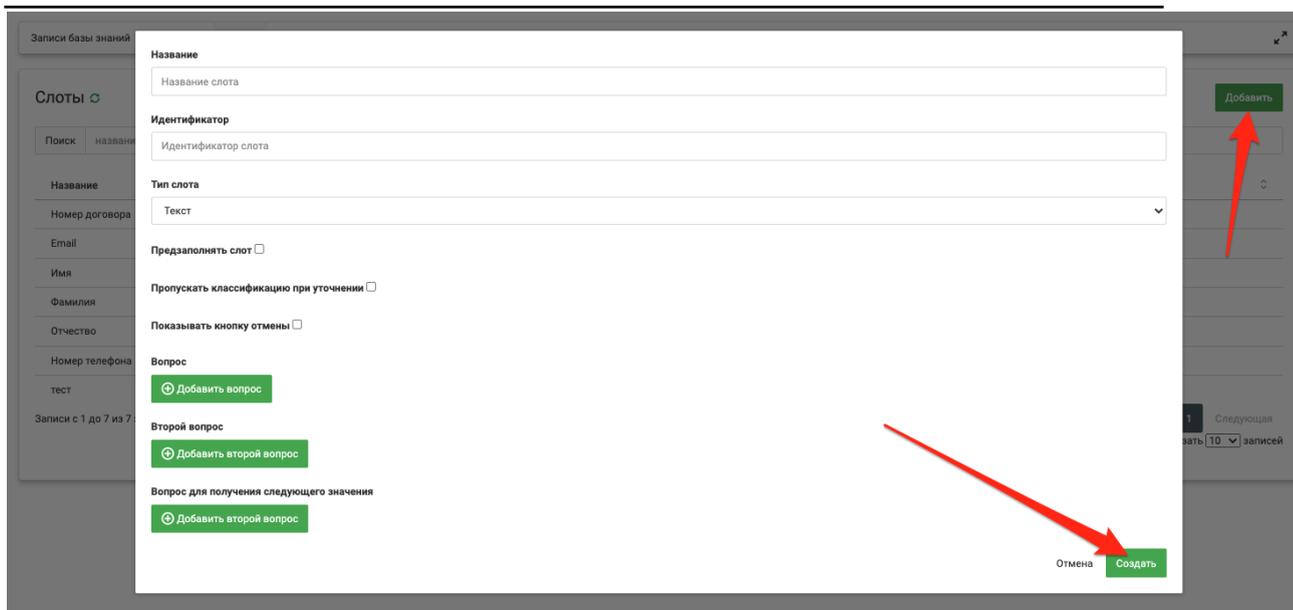


Рисунок 3.76

Для входа в интерфейс редактирования слота (рис. 3.77) необходимо кликнуть на его названии в списке слотов.

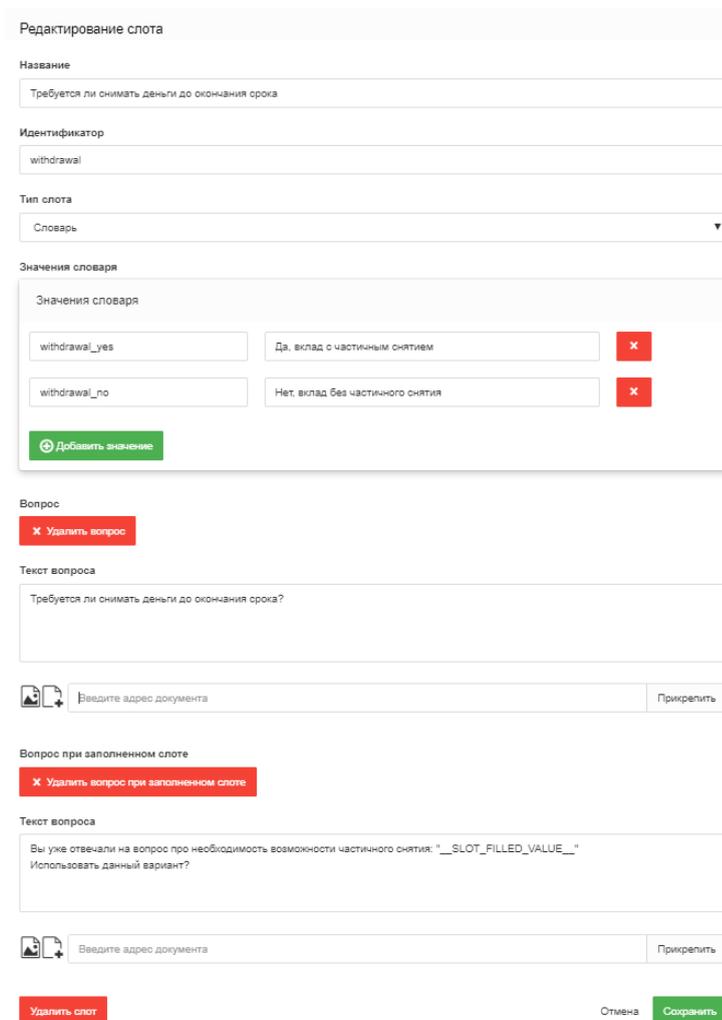


Рисунок 3.77

Тип слота выбирается из всплывающего списка (рис. 3.78).

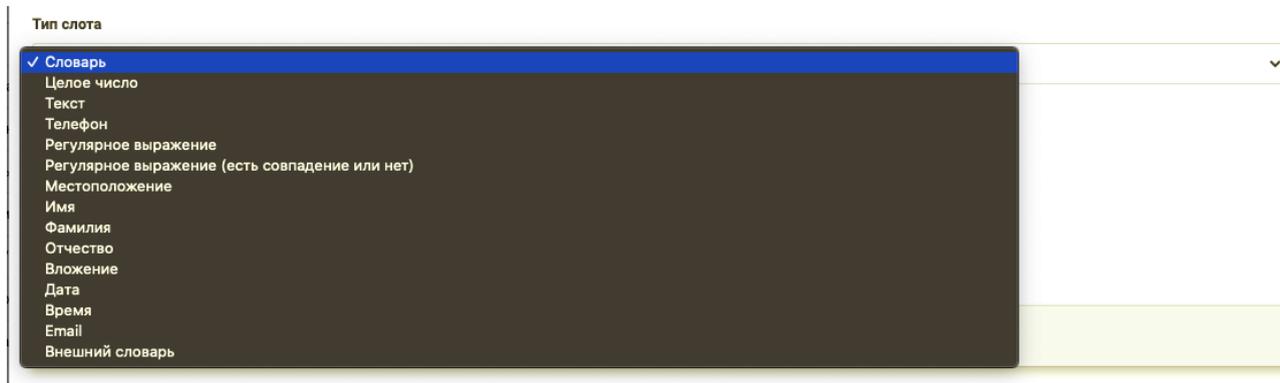


Рисунок 3.78

При создании нового слота после заполнения всех строк редактора нажать «Создать», после редактирования «Сохранить». После появления надписи о создании/сохранении слота он становится доступен в Базе знаний.

3.20.1 Слот типа словарь. Функционал синонимов.

В слоте типа "Словарь" указывается поле и синонимы к нему.

Синонимы берутся из слотов. Берутся все. Для каждого предложения и каждой группы синонимов, проверяем, есть ли вхождение одного из синонимов в предложение. Если есть, то добавляем в обучающую выборку примеры, где в предложении найденное вхождение заменено каждым синонимом.

Поиск слова в предложении нечеткий, как и у слотов. Граница уверенности регулируется через `slot_threshold`:

0 - заменять все.

1 - заменять только точное совпадение.

Значение по умолчанию - 0.9.

Инструкция по добавлению функционала синонимов:

1. Необходимо в Базе знаний в разделе Слоты создать Слот типа “Словарь”.

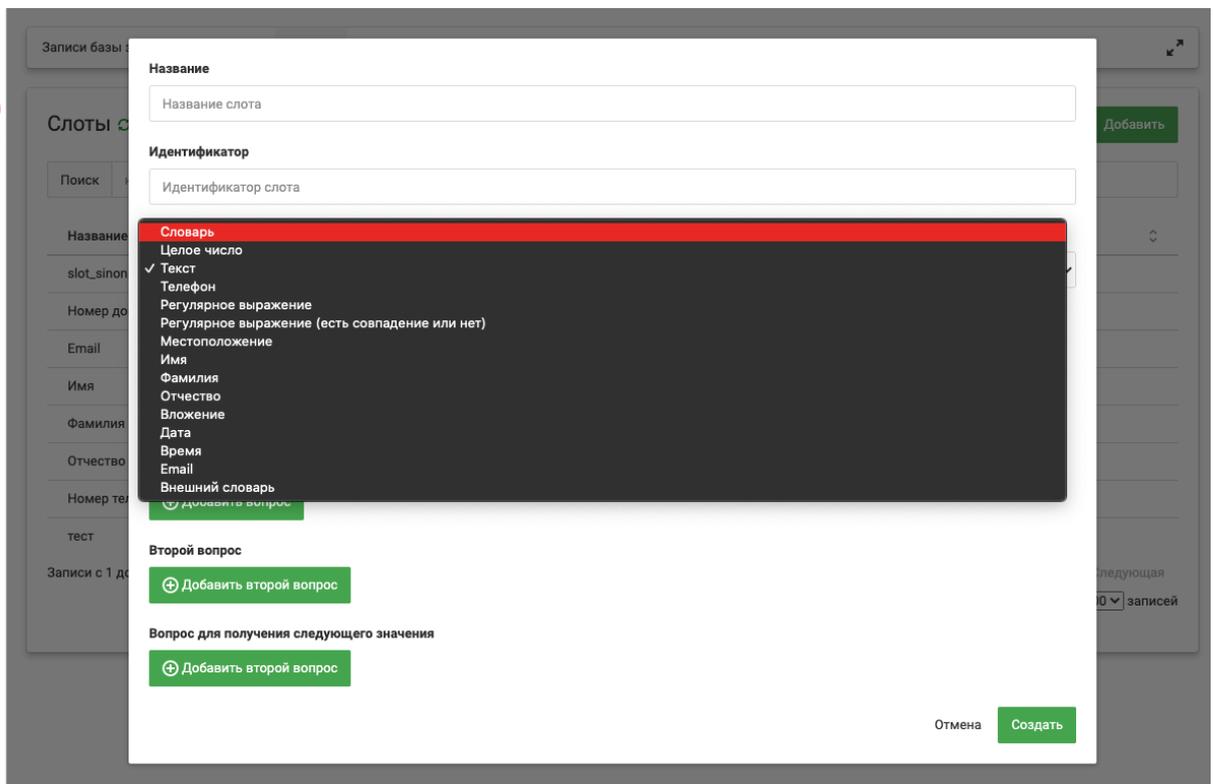


Рисунок 3.79

2. Указываем любой идентификатор и запоминаем его.

В качестве значений Словаря используется Ключевое слово и в графу “Синонимы” вносим слова-синонимы в данному слову.

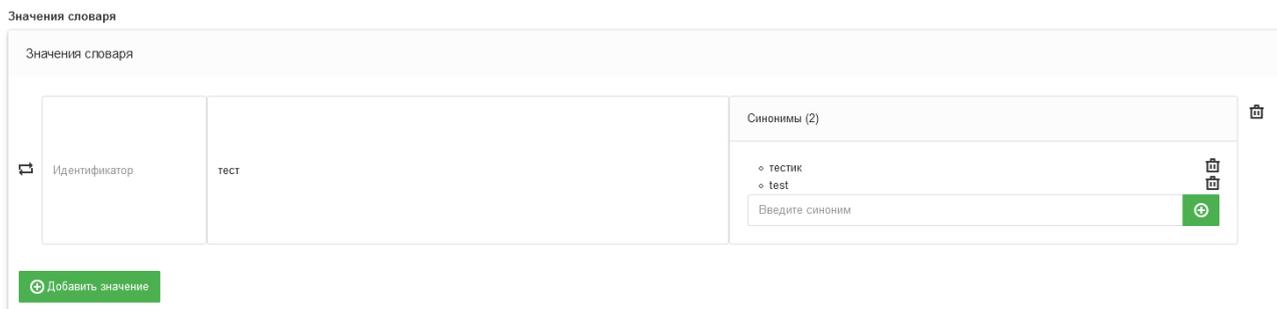


Рисунок 3.80

3. Сохраняем слот. Переходим к настройке классификаторов. База знаний, раздел - Управление классификаторами.



Рисунок 3.81

4. В настройках классификатора в поле “Делать аугментацию синонимами” ставим галочку.

В поле “Конфиг индекса слотов (обязательно при аугментации синонимами)” указываем строку следующего вида:
`http://<host>:<port>/<index_name>/<type>`

Пример (`http://elasticsearch-elasticsearch-es-http.elastic:9200/sodfu_slot/slot`)

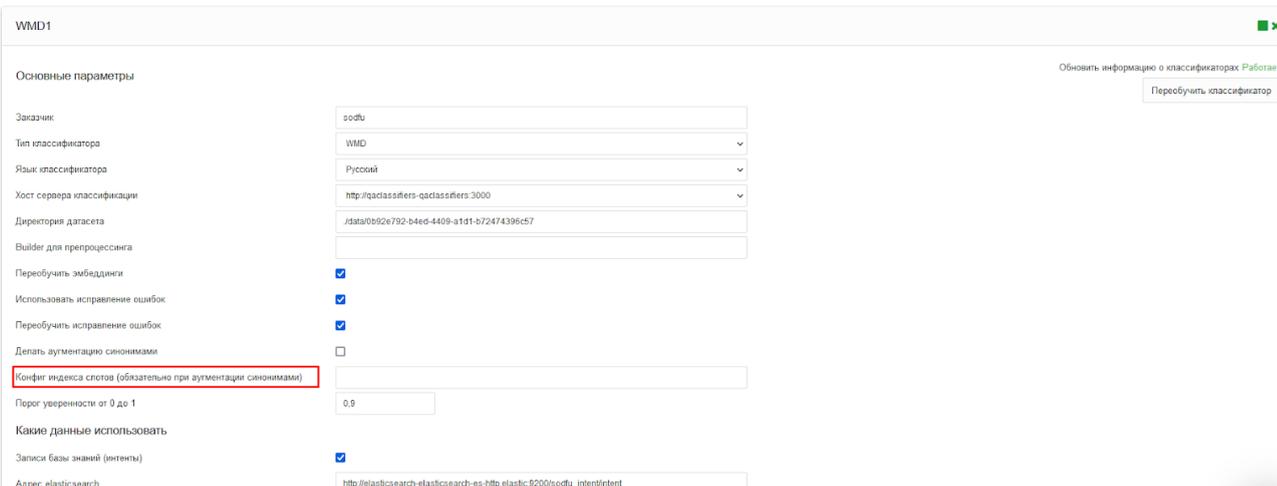


Рисунок 3.82

5. Далее необходимо указать “Порог уверенности от 0 до 1” - поиск слова в предложении нечеткий, как и у слотов. Граница уверенности регулируется через `slot_threshold`: 0 - заменять все. 1 - заменять только точное совпадение. Дефолт - 0.9.

3.20.2 Аргументация синонимами

Эта функция позволяет расширить обучающую выборку, чтобы классификаторы допускали меньше ошибок.

При включении этой функции генерируются новые примеры для обучения на основе тех, что есть в примерах, заведённых вручную или размеченных оператором.

Алгоритм простой: заменяем слова в сообщении на их синонимы. Синонимы берутся из слотов с типом «Словарь»

Как работает

Синонимы берутся из слотов. Берутся все. Для каждого предложения и каждой группы синонимов, проверяем, есть ли вхождение одного из синонимов в предложение. Если есть, то добавляем в обучающую выборку примеры, где в предложение найденное вхождение заменено каждым синонимом.

Поиск слова в предложении нечеткий, как и у слотов. Граница уверенности регулируется через `slot_threshold`: 0 - заменять все. 1 - заменять только точное совпадение. Дефолт - 0.9.

Техническая информация:

Новые поля в конфиге эластика, что в конфиге датасета:

`do_synonyms_augmentation` (bool) - делать аугментацию синонимами или нет. default - False

`slot_index` - конфиг индекса слотов (так же как и все остальные индексы). Отсюда будут браться синонимы. Обязательно, если `do_synonyms_augmentation` = True

`slot_threshold` (float) - порог уверенности от 0 до 1, принимается найденное вхождение, как слово для замены синонимами или нет. 0 - заменять все. 1 - заменять только точное совпадение. Дефолт - 0.9

пример полного конфига датасета:

```
{
  "customer_id": "rupost",
  "folder": "./data/rupost",
  "topicFile": "topics.txt",
  "answerFile": "answers.txt",
  "questionFile": "questions.txt",
  "intentFile": "intents.txt",
  "embeddingsModel": "embeddings",
  "spellCorrectionModel": "spell",
  "preprocessorBuilder": "preprocessor_builders.post",
  "lang": ["ru"],
  "elasticsearch": {
    "keep_old_dataset": false,
    "use_intent_qa_id": true,
    "check_intents_against_topics": false,
    "indices": [
      {"host": "<http://localhost>",
        "index_name": "rupost_intent",
        "port": 9200,
        "filters": [
          {"type": "intent",
            "fields": [{"remove_if_absent": false, "value": "st", "field_name":
"record_type"}]}
        ]},
      {"types": ["intent"]}
    ], {
      "host": "<http://localhost>",
```

```
"index_name": "rupost_message",
  "port": 9200,
  "types": ["message"]
}
],
"use_intent_kb_id": false,
"download_dump": true,
"check_questions_against_answers": false,
"do_update": true,
"remove_dump_files_after_export": true,
  "do_synonyms_augmentation": true,
  "slot_index": {
    "host": "<http://localhost>",
    "index_name": "rupost_slot",
    "port": 9200,
    "types": ["slot"]
  },
  "slot_threshold": 0.9
}
}
```

3.21 Вкладка «Стикер»

Раздел «Стикер» позволяет использовать стикеры в ответах. Для использования стикеров в ответах необходимо внести их в Базу знаний, для этого выбрать раздел «Стикер» (рис. 3.83).

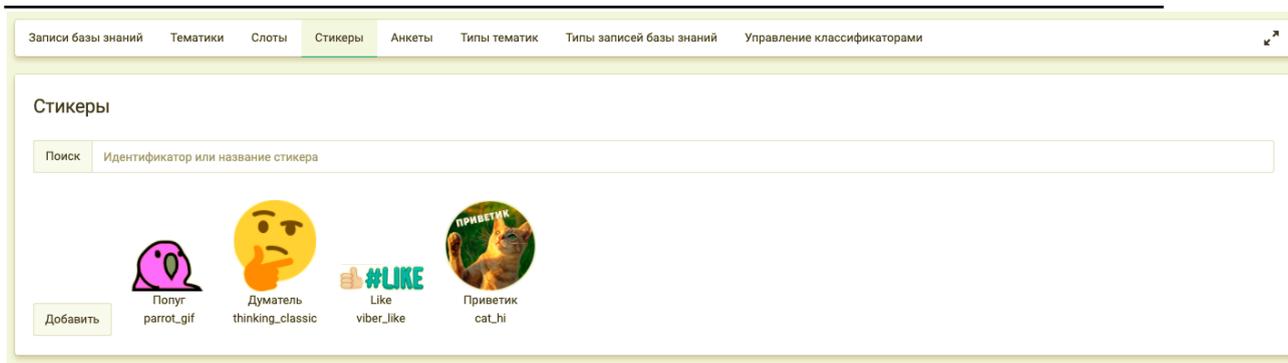


Рисунок 3.83

В разделе возможно внесение новых стикеров и поиск их по названию.

Для добавления стикера необходимо нажать кнопку «Добавить». Затем в появившейся форме (рис. 3.84) заполнить строки внося название, идентификатор и адрес и ссылку на нужный стикер.

Стикеры можно внести из каналов: Viber, Telegram, ОК или из Webchata внося соответствующий идентификатор стикера в канале.

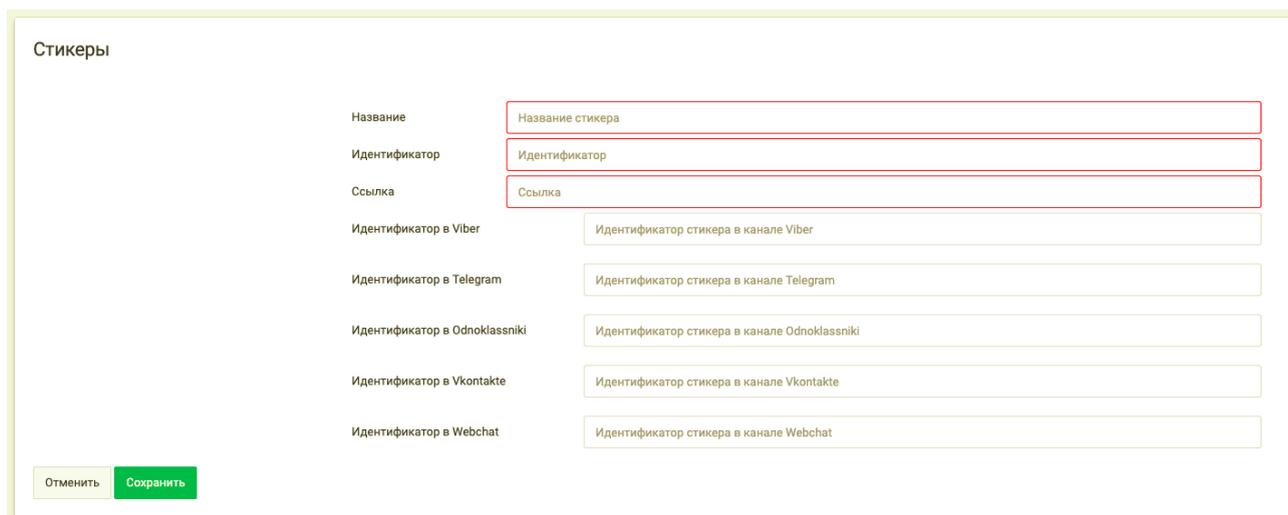


Рисунок 3.84

3.21.1 Прием стикеров

В системе КТ поддерживается прием стикеров из следующих мессенджеров:

- Facebook.
- Odnoklassniki.
- Telegram.
- Viber.
- VKontakte.

Для стикеров, заведенных в систему, изображение стикера будет видно оператору. Для остальных стикеров, отображение оператору зависит от возможностей канала и в общем случае невозможно

3.21.2 Отправка стикеров

В системе КТ поддерживаются отправка стикеров по следующим каналам:

- Facebook (отправка картинкой).
- Odnoklassniki.
- Telegram.
- Viber.
- VKontakte (отправка картинкой).

Для того, чтобы бот мог отправлять стикеры, необходимо сначала завести стикеры в систему.

3.21.3 Добавление стикера для канала Telegram

Для добавления стикера в канал Telegram необходимо последовательно выполнить действия:

1. Иницируем чат и присылаем стикер.
2. Копируем ссылку на стикер, кликая правой кнопкой на стикер и копируя её (рис. 3.85).



Рисунок 3.85

Для примера это:

```
http://localhost:8100/file//024/142e6e81-d042-445e-850e-0ed5863305f7-28aee673-b0a7-40fa-b7f1-59eec8b551e5.webp
```

3. Копируем идентификатор стикера в канале Telegram открыв инструменты разработчика и скопировав значение data-sticker-id на элементе изображения стикера (рис. 3.86).

```
 == $0
```

Рисунок 3.86

Для примера это:

CAACAgIAAxkBAAM8XxcJu5vjqtPb039jnp4qHxSBRhCAAUA8LvXAXbEALQnEa-sxoE

4. Создаем новый стикер в разделе База знаний – Стикеры (рис. 3.87).

5. Готово. Теперь созданный стикер можно использовать в статьях базы знаний в markdown.

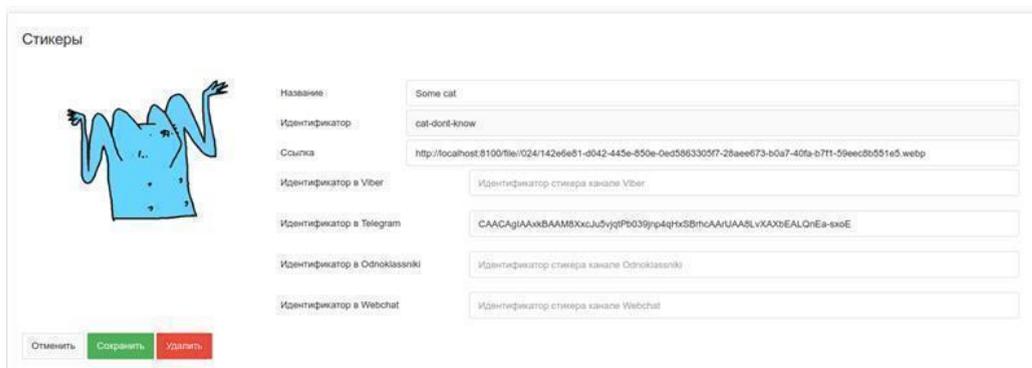


Рисунок 3.87

Для того чтобы добавить отправку этого стикера в другие каналы, можно также скопировать идентификатор этого или аналогичного стикера и подставить его значение в созданный стикер.

3.21.4 Добавление стикера для канала ВКонтакте

Последовательно выполнить действия:

1. Инициуем чат и присылаем стикер.
2. Копируем ссылку на стикер.
3. Создаем новый стикер в разделе База знаний - Стикеры (ничего не добавлять в поле Идентификатор в Telegram).

3.21.5 Добавление стикера для канала Одноклассники

Последовательно выполнить действия:

1. Инициуем чат и присылаем стикер.
2. Копируем ссылку на стикер.
3. Копируем идентификатор стикера в канале, открыв инструменты разработчика и скопировав значение data-sticker-id на элементе изображения стикера (`src="https://i.mycdn.me/getSmile?smileId=d28ee61700&id` стикера это все что после smileId= и до &).
4. Создаем новый стикер в разделе База знаний - Стикеры (ничего не добавлять в поле Идентификатор в Telegram, добавить в поле идентификатор в одноклассниках id стикера, полученного на предыдущем шаге).

3.21.6 Добавление стикера для канала Facebook

Последовательно выполнить действия:

1. Инициуем чат и присылаем стикер.
2. Копируем ссылку на стикер.
3. Создаем новый стикер в разделе База знаний – Стикеры (ничего не добавлять в поле Идентификатор в Telegram).
4. Добавить ссылку на изображение для стикера. Это изображение будет использоваться для отправки стикера в каналы, где стикеры не поддерживаются. Для получения ссылки можно загрузить изображение в статью базы знаний и скопировать ссылку на загруженное изображение.

Обязательно проставлено поле "Вебхук для загрузки файлов" у канала Facebook в настройках.

3.22 Вкладка «Анкеты»

Во вкладке создаётся форма хранения информации анкеты Обратившегося (рис. 3.88).

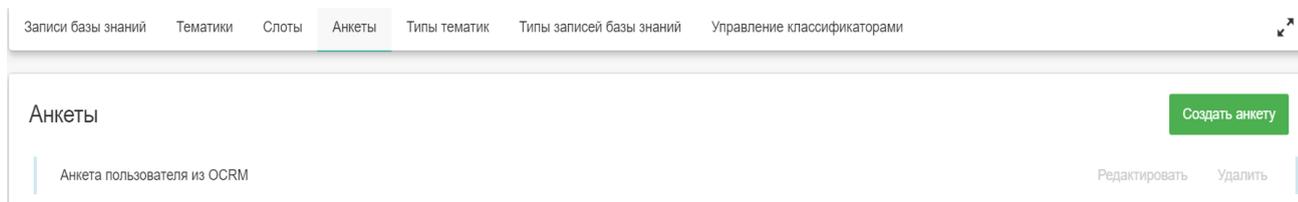


Рисунок 3.88

Поля анкеты можно создать и отредактировать во вкладке «Редактирование анкеты» (рис. 3.89), нажав на названии анкеты, в «Общая информация». Для просмотра и редактирования полей анкеты нажать на названии анкеты.

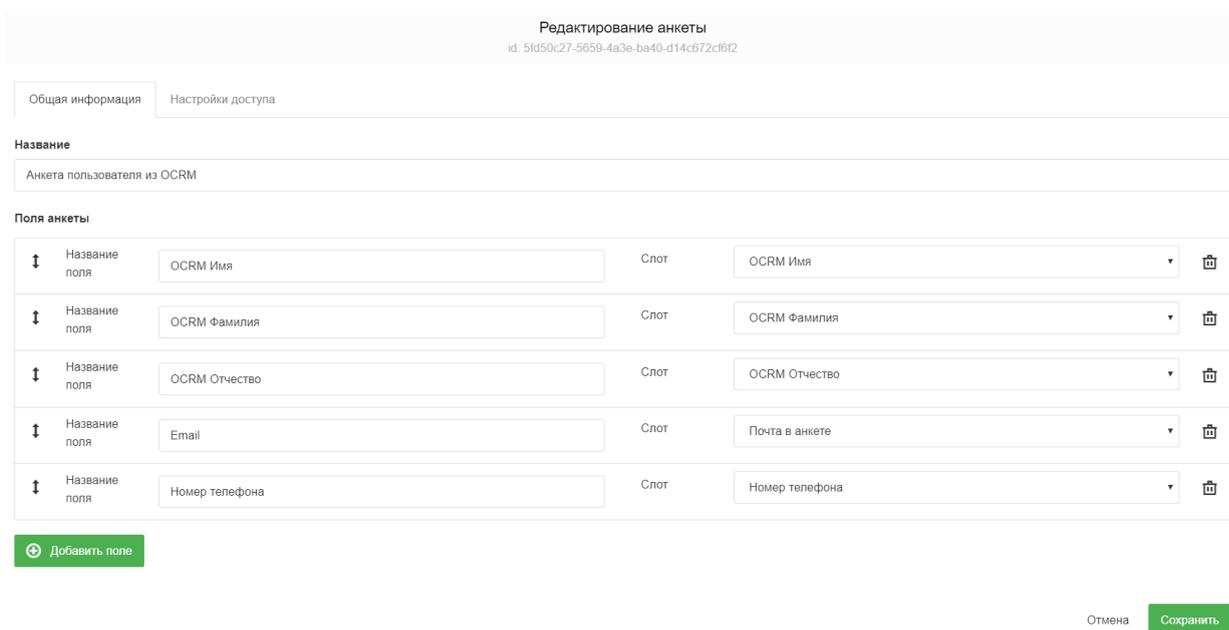


Рисунок 3.89

Анкету можно создать, нажав «Создать Анкету» и отредактировать, нажав «Редактировать» (рис. 3.90).

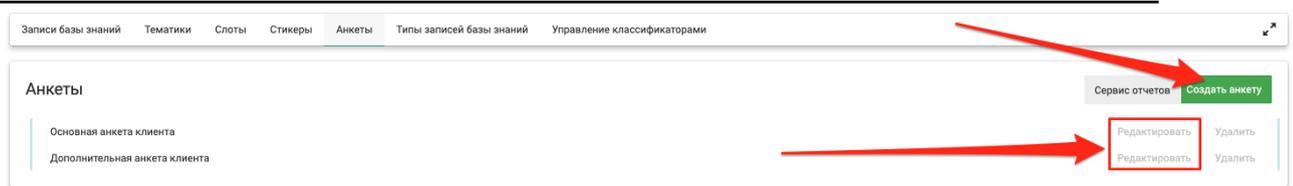


Рисунок 3.90

Для добавления поля анкеты нажать «Добавить поле» и в появившейся форме заполнить поля (рис. 3.91).

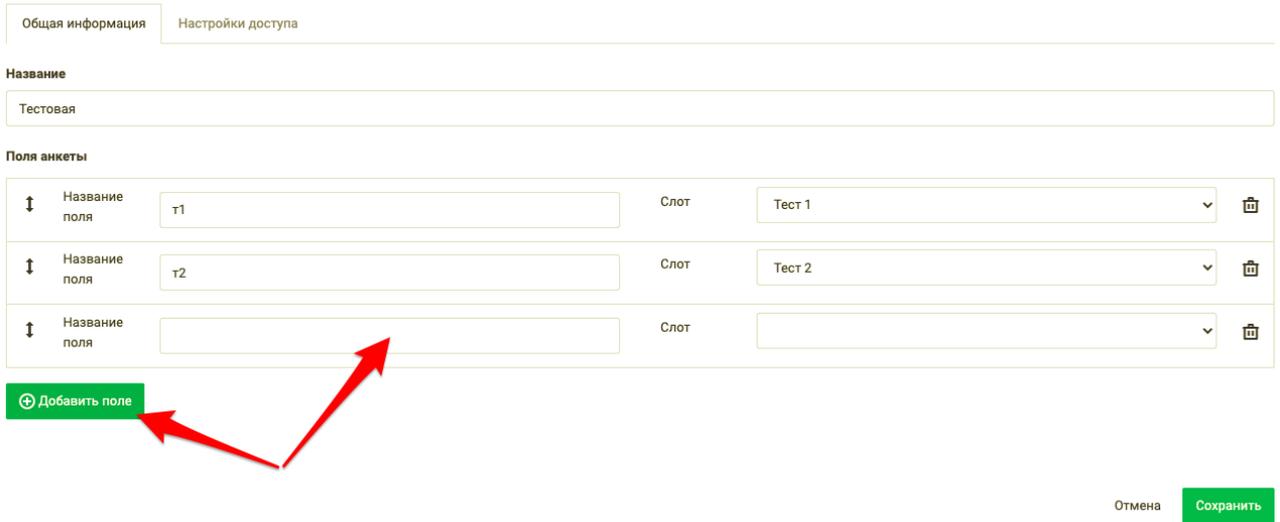


Рисунок 3.91

Слот выбирается из всплывающего списка после нажатия . Сохранение нового поля осуществляется нажатием «Создать».

Во вкладке «Настройка доступа» можно установить линию, с которой эта анкета будет доступна (рис. 3.92).

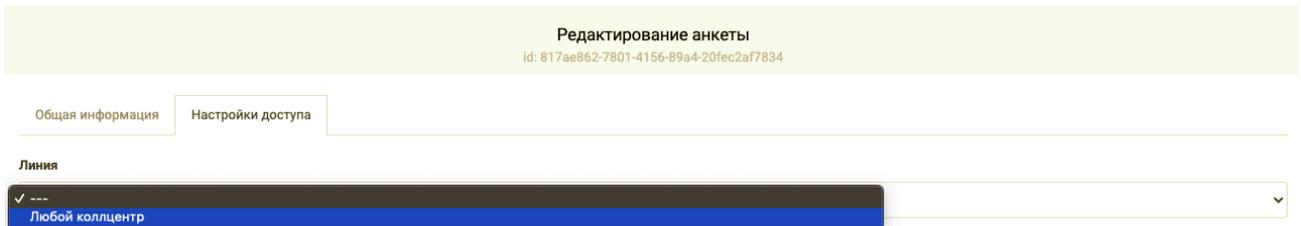


Рисунок 3.92

3.23 Описание классификаторов

На платформе может быть настроено неограниченное количество классификаторов. В системе представлено 4 основных вида классификаторов:

- customer_LOGREG (тип - LOGREG).
- customer_GRU (тип - GRU).
- customer_KNN (тип – KNN).
- customer_WMD (тип - WMD).

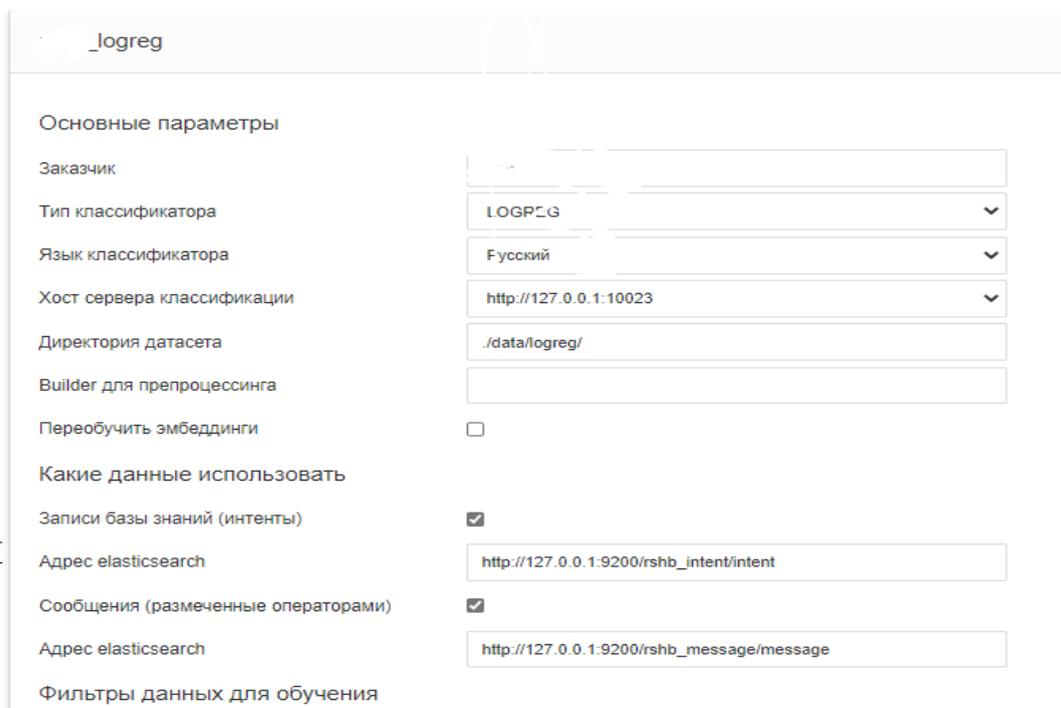
Для классификации каждого запроса одновременно используются все 4 классификатора. Для выдачи ответа используется результат классификатора, который вернул наиболее высокую оценку точности классификации (score) для ответа. Для каждого классификатора имеется возможность указать порог отсечения по оценке точности.

В целом классификаторы типа WMD позволяют отвечать на конкретные вопросы, достаточно близкие по формулировкам примерам в базе знаний.

Классификаторы типов LOGREG и GRU позволяют обрабатывать более сложные реальные формулировки, основываясь на большом количестве примеров, но не гарантируют ответа на конкретный вопрос, даже если он очень близок к примеру в БЗ (если такая формулировка единичная и не поддерживается в целом другими примерами). Как правило, именно эти классификаторы снимают основной объем трафика по часто используемым запросам.

3.23.1 Классификатор Customer _LOGREG

Классификатор, основанный на признаках, извлекаемых из текста. Данный классификатор позволяет выявлять зависимости в вопросах пользователей, на основании которых можно выбирать необходимую статью БЗ (рис. 3.93).



| _logreg | |
|-------------------------------------|--|
| Основные параметры | |
| Заказчик | |
| Тип классификатора | LOGREG |
| Язык классификатора | Русский |
| Хост сервера классификации | http://127.0.0.1:10023 |
| Директория датасета | /data/logreg/ |
| Builder для препроцессинга | |
| Переобучить эмбединги | <input type="checkbox"/> |
| Какие данные использовать | |
| Записи базы знаний (интененты) | <input checked="" type="checkbox"/> |
| Адрес elasticsearch | http://127.0.0.1:9200/rshb_intent/intent |
| Сообщения (размеченные операторами) | <input checked="" type="checkbox"/> |
| Адрес elasticsearch | http://127.0.0.1:9200/rshb_message/message |
| Фильтры данных для обучения | |

Рисунок 3.93

Производительность классификатора не зависит от количества примеров, на которых обучен данный классификатор, поэтому его можно обучать на полном объеме примеров (то есть с ростом примеров скорость работы не изменяется).

Эффективность классификатора возрастает с количеством примеров. При количестве примеров более 40 достигается высокая эффективность.

Обучается существенно быстрее (порядка 1 часа на 100 000 примеров на CPU, не может использовать GPU). Менее требователен к количеству примеров (рекомендуется от 10 примеров на интент). Скорость классификации не зависит от количества примеров и интентов.

Примечание: добавление одиночного примера для данного классификатора не гарантирует классификацию в том числе для 100% совпадающего вопроса.

3.23.2 Классификатор Customer_GRU

Deep learning классификатор, использующий рекуррентную сеть с механизмом внимания. Данный классификатор производит лемматизацию слов в запросе, с переводом слов в эмбединги (дообучаемые на информации из базы знаний, что позволяет эффективно использовать специфические термины предметной области) (рис. 3.94).

The screenshot displays the configuration page for the 'customer' classifier. The main settings include:

- Основные параметры:**
 - Заказчик: customer
 - Тип классификатора: CNN
 - Хост сервера классификации: http://192.168.100.100:10025
 - Директория датасета: /data/test/
 - Builder для препроцессинга: (empty)
- Какие данные использовать:**
 - Записи базы знаний (интенты):
 - Адрес elasticsearch: http://localhost:20836/rshb_intent/intent
 - Сообщения (размеченные операторами):
 - Адрес elasticsearch: (empty)
- Фильтры данных для обучения:**
 - Фильтр: Для каких данных: Записи базы знаний (интенты)
 - Тип фильтра: По количеству примеров на намерение
 - Настройка фильтра: Больше, чем: 10; Больше (включая), чем: (empty); Меньше, чем: (empty); Меньше (включая), чем: (empty)

Additional UI elements include a 'Текущий статус классификатора: Работает' indicator and a 'Переобучить классификатор' button.

Рисунок 3.94

Данный классификатор использует глубокую нейронную сеть для определения подходящей статьи базы знаний и позволяет наиболее качественно определить подходящую статью базы знаний для различных формулировок вопросов пользователей.

Это наиболее эффективный классификатор в случае наличия большого количества разметки (от 25 до десятков тысяч примеров на интент).

Скорость обучения классификатора сильно зависит от количества примеров (порядка 10 часов на 100 000 примеров на CPU, может быть ускорен за счет использования GPU). Скорость классификации не зависит от количества примеров и интентов.

Эффективность классификатора возрастает с количеством примеров. При количестве примеров более 50-100 достигается высокая эффективность

Производительность классификатора не зависит от количества примеров, на которых обучен данный классификатор, поэтому его можно обучать на полном объеме примеров (то есть с ростом примеров скорость работы не изменяется).

Примечание: добавление одиночного примера для данного классификатора не гарантирует классификацию в том числе для 100% совпадающего вопроса.

3.23.3 Классификатор Customer_KNN

Классификатор, подбирающий близкий пример из обучающей выборки и оценивающий степень соответствия. Для своей работы использует результаты обучения deep learning классификатора. Эффективен для интентов с любым количеством примеров. Скорость классификации зависит от количества примеров, но используемый алгоритм позволяет эффективно работать с несколькими сотнями тысяч примеров. Рекомендуется при наличии "широкой" базы интентов, среди которых много интентов с малым количеством разметки.

3.23.4 Классификатор customer_WMD

Классификатор находит статью БЗ, где имеется близкий к вопросу пользователя пример запроса. При сравнении допускается добавление незначимых слов, перестановки слов, использование синонимов и слов в других роде и падеже.

Классификатор, подбирающий близкий пример из обучающей выборки и оценивающий степень соответствия (рис. 3.95).

Основные параметры

Текущий статус классификатора: Работает

Переобучить классификатор

Заказчик

Тип классификатора: WMD

Язык классификатора: Русский

Хост сервера классификации: http://127.0.0.1:10023

Директория датасета: /data/wmd/

Builder для препроцессинга

Переобучить эмбединги

Какие данные использовать

Записи базы знаний (интенты)

Адрес elasticsearch: http://127.0.0.1:9200/rshb_intent/intent

Сообщения (размеченные операторами)

Адрес elasticsearch

Фильтры данных для обучения

Фильтр

Для каких данных: Записи базы знаний (интенты)

Тип фильтра: По количеству примеров на намерение

Фильтр по количеству примеров на намерение (интент)

Больше, чем:

Меньше, чем:

Рисунок 3.95

Классификатор эффективен для обеспечения гарантированного ответа на вопросы, для которых близкие примеры в явном виде присутствуют в базе знаний.

Скорость классификации зависит от количества примеров, но используемый алгоритм позволяет эффективно работать с несколькими сотнями тысяч примеров. Особенностью данного классификатора является то, что скорость классификации запроса пользователя снижается при увеличении количества примеров, поэтому рекомендуется, чтобы у данного классификатора в сумме было не более 1000 примеров для сравнения.

Рекомендуется при наличии "широкой" базы интентов, среди которых много интентов с малым количеством разметки.

По этой причине данный классификатор настроен для работы только со статьями базы знаний, где имеется небольшое количество примеров вопросов пользователей (395 менее 50). Для статей с большим количеством примеров (реальных запросов пользователя с близкими формулировками одного и того же

запроса) необходимо по причине производительности и качества классификации использовать другие типы классификаторов.

Переобучение данного классификатора осуществляется быстро - порядка 5-10 минут.

Для проверки работы классификатора можно использовать примеры вопросов идентичные или близкие к указанным в статьях примерам. Данный классификатор не работает со статьями с более чем 1000 примерами.

3.23.5 Рекомендации по переобучению классификаторов

Классификаторы WMD рекомендуется обучать по мере возникновения такой потребности (то есть при изменении примеров вопросов в соответствующих статьях БЗ и необходимости применить измененные настройки). Пример обучения классификатора WMD приведен в Приложении 1.

Классификатор LOGREG рекомендуется обучать по мере возникновения такой потребности. Необходимо учитывать, что единичные изменения примеров скорее всего не скажутся на результатах классификации, поэтому при таких изменениях смысла переобучать классификатор нет.

Классификатор GRU рекомендуется обучать в ночное время по мере накоплении существенного количества изменений в разметке примеров.

При внесении примеров не требуется никакой дополнительной предобработки текста вопросов пользователей. Вопросы должны быть в той форме, как их задают пользователи.

В идеале число примеров на статью БЗ должно быть пропорционально числу вопросов по этой статье в реальной эксплуатации. Это автоматически достигается при разметке максимального объема задаваемых пользователями вопросов операторами (при ответе статьёй из базы знаний, используя “самолетик”) или при разметке истории диалогов. В целом такая разметка реальных вопросов клиентов – это основной механизм, позволяющий эффективно снимать трафик с операторов.

3.34.6 Сбор разметки для классификаторов

Разметка для обучения классификатор накапливается в базе знаний. Платформа предоставляет несколько способов сбора разметки и последующего управления разметкой:

- Автоматическое накопление разметки для классификаторов в процессе работы операторов и использования ими статей Базы знаний для ответов.

- Инструмент разметки истории диалогов.

- Инструменты подбора разметки для интента на основе неразмеченной истории диалогов (полуавтоматически на базе имеющихся примеров и с помощью ключевых слов).

Все изменения, вносимые во вкладке «Записи базы знаний», приводят к необходимости переобучения Бота и Электронного виртуального помощника (ЭВА).

Переобучение происходит по расписанию (обычно ночь) либо возможно его произвести вручную.

Для запуска обучения необходимо нажать на кнопку «Обновить информацию о классификаторах». Далее на вкладке найти строку `customer_wmd` и выполнить процедуру переобучения классификаторов описанную ниже (см. п.3.24).

3.24 Вкладка «Управление классификаторами»

Все изменения, вносимые во вкладке «Записи базы знаний», приводят к необходимости переобучения Бота и Электронного виртуального помощника (ЭВА).

Переобучение происходит по расписанию (обычно ночью) либо возможно его произвести вручную.

Переобучение классификаторов осуществляется на странице «База знаний» / «Управление классификаторами» (рис. 3.96):

- В разделе формы «customer_классификатор, например customer-knp» (если классификаторы не видны нажать «Обновить информацию о классификаторах») справа нажать «Переобучить классификатор».

- Дождаться, когда надпись «Работает» изменится на красную надпись «Обучается», а потом снова станет зелёной (классификатор переобучился).

Классификаторы

Обновить информацию о классификаторах Загрузить сохраненную копию

frontline_knn

Текущий статус классификатора: **Работает**

Основные параметры

| | | |
|-------------------------------------|---|--|
| Заказчик | <input type="text" value="customer"/> | <input type="button" value="Переобучить классификатор"/> |
| Тип классификатора | <input type="text" value="KNN"/> | |
| Язык классификатора | <input type="text" value="Русский"/> | |
| Хост сервера классификации | <input type="text" value="http://10.10.60.56:10030"/> | |
| Директория датасета | <input type="text" value="/data/knn/"/> | |
| Builder для препроцессинга | <input type="text"/> | |
| Переобучить эмбединги | <input type="checkbox"/> | |
| Какие данные использовать | | Настройки изменены, нажмите сюда чтобы переобучить |
| Записи базы знаний (интенты) | <input checked="" type="checkbox"/> | |
| Адрес elasticsearch | <input type="text" value="http://10.10.60.63:10092/frontline_intenti/intent"/> | |
| Сообщения (размеченные операторами) | <input checked="" type="checkbox"/> | |
| Адрес elasticsearch | <input type="text" value="http://10.10.60.63:10092/frontline_message/message"/> | |
| Фильтры данных для обучения | | |
| Фильтры отсутствуют | <input type="button" value="Добавить фильтр"/> | |

Рисунок 3.96

Не рекомендуется менять настройки классификаторов без предварительной консультации с технической поддержкой.

Внимание! Изменение содержимого статей не требует переобучения классификаторов и становится доступно клиентам немедленно после публикации статьи. Переобучение классификаторов необходимо в случае добавления новой разметки (она появляется в статье на вкладке “Размеченные сообщения”), новых примеров вопросов в статье, удалении, деактивации статей, которые содержали примеры вопросов или разметку, добавлении новых статей с примерами вопросов, а также при внесении изменений в типы записей, если эти типы задействованы в фильтрах классификаторов.

Для переобучения классификатора требуется нажать кнопку “переобучить классификатор” один раз. Изменение статуса на “переобучение” может занимать длительное время и для быстро обучаемых классификаторов может по факту оказаться незаметным для пользователя (это не является ошибкой).

При переобучении классификатора старая версия классификатора продолжает работать до момента окончания обучения и переключения на вновь обученную версию модели.

Обучение классификаторов ответственные пользователи могут осуществлять самостоятельно.

3.25 Смена привязки к статье БЗ и Обучение Бота

Бот умеет распознавать намерения Обратившегося и выбирать необходимую статью БЗ (при наличии релевантной статьи) для обработки запроса Обратившегося. Статья БЗ может содержать или готовый ответ (один или несколько) или сценарий (в этом случае бот может дозапрашивать информацию и организовывать диалог).

Чат-бот (агент AIAssist2) предоставляет несколько способов для ветвления диалога, конфигурируемые непосредственно в базе знаний:

- кнопки в ответах (в том числе, неотображаемые и срабатывающие в рамках "гиперлокальной" классификации). Такая кнопка по сути является ссылкой на другую статью базы знаний
- ветвление на базе значений слотов. В рамках одной статьи базы знаний можно настроить несколько ответов, каждый из которых выдается при выполнении определенных условий на значения слотов. Пример – разные ответы в зависимости от канала обращения пользователя, который записан в системном слоте `channel_id`.

Для эффективного накопления знаний Ботом, его более быстрого обучения Операторы в своей работе должны пользоваться статьями Базы знаний при ответах Обратившемуся, т.е. крайне важно как можно чаще при общении Оператора с Обратившимся использовать кнопку отправки готового ответа с иконкой  ("самолетик").

Для обучения Бот использует примеры фраз и реальных вопросов Обратившихся, которые получает:

- из статей Базы знаний (раздел «Пример вопросов»);
- автоматически при работе Оператора, если Оператор использует для ответов статьи БЗ (ответ с помощью );
- при ручной разметке сообщений в истории обращений (главное меню "История обращений" -> функция "Автоответ").

При необходимости подкорректировать выбор Оператора можно произвести ручное изменение выбора ответа Оператора в истории завершенных обращений.

Алгоритм ручной разметки истории обращений (главное меню «История обращений»):

- Найти нужное обращение в истории либо выбрать все обращения в статусе «ЗАВЕРШЕН» с помощью фильтра.
- Для нужных реплик пользователя выбрать соответствующую статью Базы знаний (иконка «карандаш» рядом с автоответом **Автоответ:** ). Для этого надо ввести название статьи и после появления подсказок выбрать нажав нужную статью.
- Зафиксировать свой выбор путем нажатия на кнопку “Изменить” в окне выбора статьи БЗ (рис. 3.97).
- Далее в главном меню «База знаний» открыть «Управление классификаторами» и выполнить процедуру обучения классификаторов.

Выберите статью БЗ

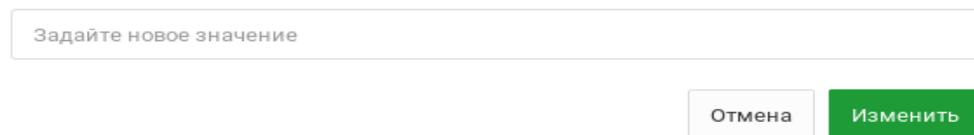


Рисунок 3.97

После этого обращению можно поставить статус «ПРОВЕРЕН», чтобы отметить статус обработки (разметки) обращения (рис. 3.98).



Рисунок 3.98

Аналогично разметку можно проводить и на вкладке "Предыдущие обращения" в меню "Поиск обращений".

4 Сценарии

Платформа ориентирована в первую очередь на использование машинного обучения и конфигурирование сценариев с использованием базы знаний, а также на накопление в базе знаний разметки на основе взаимодействия с оператором для дообучения чат-бота, и позволяет подключение внешних интеграций с помощью механизма агентов.

4.1 Возможности Чат-бота

Платформа CraftTalk предоставляет следующие встроенные возможности для создания чат-ботов для автоматизации обработки запросов:

1. Распознавание намерения пользователя (интента).
2. Организация диалога. Организация диалога с пользователем для запроса у пользователя необходимой информации с возможностью ветвления диалога в зависимости от ответов пользователя, валидации введенной информации.
3. Извлечение сущностей (доступно в AIAssist2). Автоматическое извлечение сущностей различного типа с использованием встроенных и внешних экстракторов, в том числе извлечение сущностей из исходного запроса пользователя и в режиме диалога при необходимости. Валидация введенных значений для разных типов сущностей. Экстрактор для нечеткого извлечения значений по словарям с возможностью использования множественных синонимов для каждого значения.
4. Работа с кнопками. Вне зависимости от возможности отображения кнопок в определенном канале, обеспечивается распознавание “нажатия” кнопок в виде текстовых ответов пользователей на основе нечеткого извлечения значений словарей (с возможностью использования синонимов).
5. Хранение контекста диалога и введенной пользователем информации на протяжении всего диалога.
6. Возможность подключения интеграций для обработки навыков и передачи полученной у пользователя (и обработанной) информации в интеграции. Интеграция может как осуществлять конкретные действия (например, передать информацию в какую-то систему), так и являться полноценным чат-ботом для

обработки сложных сценариев. Интеграция осуществляется с помощью механизма агентов обработки запросов.

7. Автоматическое дообучение чат-бота на основе информации по обработке запроса оператором. Нераспознанный чат-ботом запрос может быть передан оператору. Оператор на основе поиска в базе знаний или подсказки ИИ-ассистента выбирает подходящий запрос и может осуществить обратное переключение на чат-бота. Выбор оператора автоматически запоминается для дообучения чат-бота.

8. Инструмент разметки для интенгов на основе истории диалогов (накопленной в системе или загруженной из другой системы) в полуавтоматическом режиме:

8.1 Подбирать похожие вопросы на основе уже размеченных примеров.

8.2 Подбирать вопросы на основе ключевых слов.

9 Обученный smalltalk модуль, позволяющий разделять вопросы по бизнес-сценариям и вопросы на свободные темы. Позволяет для вопросов на свободные темы задать несколько контролируемых ответов, которые должны вести клиента к бизнес-сценариям.

10 Возможность использовать чат-бот в качестве ассистента для оператора-человека.

11 Поддержка любых подключенных к платформе каналов единым образом. Имеется возможность при необходимости разделения ответов в зависимости от канала.

12 Навыки (интенги) чат-бота и вся разметка для чат-бота полностью настраиваются во встроенной базе знаний и могут контролироваться заказчиком.

4.2 Распознавание сущностей и заполнение слотов

Платформа поддерживает концепцию сущностей, которые используются для заполнения слотов в процессе диалога. Слоты могут храниться в контексте диалога, пока он не завершен. Сущность – это единица информации, такая как имя, фамилия, дата, адрес электронной почты и так далее.

Поддерживаются следующие типы сущностей:

- текстовые;

- с проверкой регулярным выражением (с возможностью автоматического извлечения из запроса пользователя);
- словари (с возможностью рендеринга значений в виде кнопок и распознавания текстового ввода и автоматического извлечения значения из запроса пользователя с нечетким совпадением);
- внешние словари (с возможностью распознавания текстового ввода и автоматического извлечения из запроса пользователя с нечетким совпадением);
- числовые;
- имя, фамилия, отчество, дата, e-mail.

После распознавания интента, менеджер диалога автоматически пытается извлечь значения слотов (описанных в соответствующей статье базы знаний) из исходного запроса пользователя и при необходимости уточнить значение слотов в диалоговом режиме. В целом процесс распознавания сущностей и уточнения их значений в процессе диалога при необходимости является автоматическим, и требует только простой настройки сущности.

При необходимости более точного управления заполнением слотов, платформа позволяет возможность гибко контролировать процесс слот филлинга – подсчитывать число попыток, проверять условия после каждой попытки заполнения (таким образом, заполняя альтернативные варианты слотов, вовремя распознавания ненужные и пропуская их).

Также предоставляются команды для управления слотами: изменение слотов, обнуление слотов, заполнение слотов из значений других слотов, инкременты и т.д. Данные команды распознаются агентом AIAssist2 и могут вводиться непосредственно в ответах базы знаний.

Помимо явно конфигурируемых слотов, платформа автоматически подставляет значения некоторых системных слотов – например идентификатор канала, что позволяет использовать эту информацию для ветвления диалога, специализации ответов и для внешних агентов.

4.3 Ветвления диалога

Чат-бот (агент AIAssist2) предоставляет несколько способов для ветвления диалога, конфигурируемые непосредственно в Базе знаний:

- Кнопки в ответах (в том числе, неотображаемые и срабатывающие в рамках "гиперлокальной" классификации). Такая кнопка по сути является ссылкой на другую статью Базы знаний.

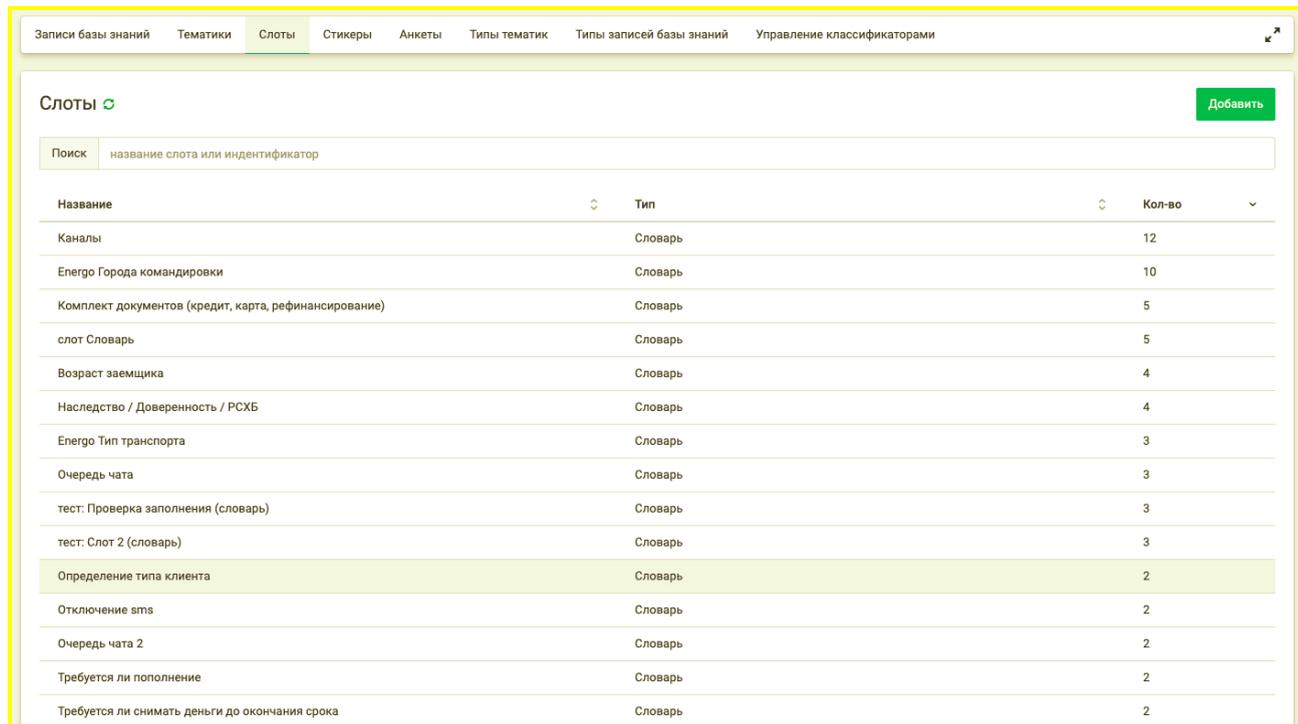
- Ветвление на базе значений слотов. В рамках одной статьи базы знаний можно настроить несколько ответов, каждый из которых выдается при выполнении определенных условий на значения слотов. Пример – разные ответы в зависимости от канала обращения пользователя, который записан в системном слоте channel_id.

4.4 Общее описание

Сценарий это описание процесса взаимодействия с пользователем с ветвлением ответов в зависимости от ответов и запросов пользователя

Первым шагом разрабатывается сценарий с последовательностью опроса и ветвлениями в зависимости от ответа.

Вторым шагом для реализации сценария сначала создаём слоты для реализации сценария и проведения опроса, например, как показано на рисунке 4.1.



| Название | Тип | Кол-во |
|---|---------|--------|
| Каналы | Словарь | 12 |
| Енерго Города командировки | Словарь | 10 |
| Комплект документов (кредит, карта, рефинансирование) | Словарь | 5 |
| слот Словарь | Словарь | 5 |
| Возраст заемщика | Словарь | 4 |
| Наследство / Доверенность / РСХБ | Словарь | 4 |
| Енерго Тип транспорта | Словарь | 3 |
| Очередь чата | Словарь | 3 |
| тест: Проверка заполнения (словарь) | Словарь | 3 |
| тест: Слот 2 (словарь) | Словарь | 3 |
| Определение типа клиента | Словарь | 2 |
| Отключение sms | Словарь | 2 |
| Очередь чата 2 | Словарь | 2 |
| Требуется ли пополнение | Словарь | 2 |
| Требуется ли снимать деньги до окончания срока | Словарь | 2 |

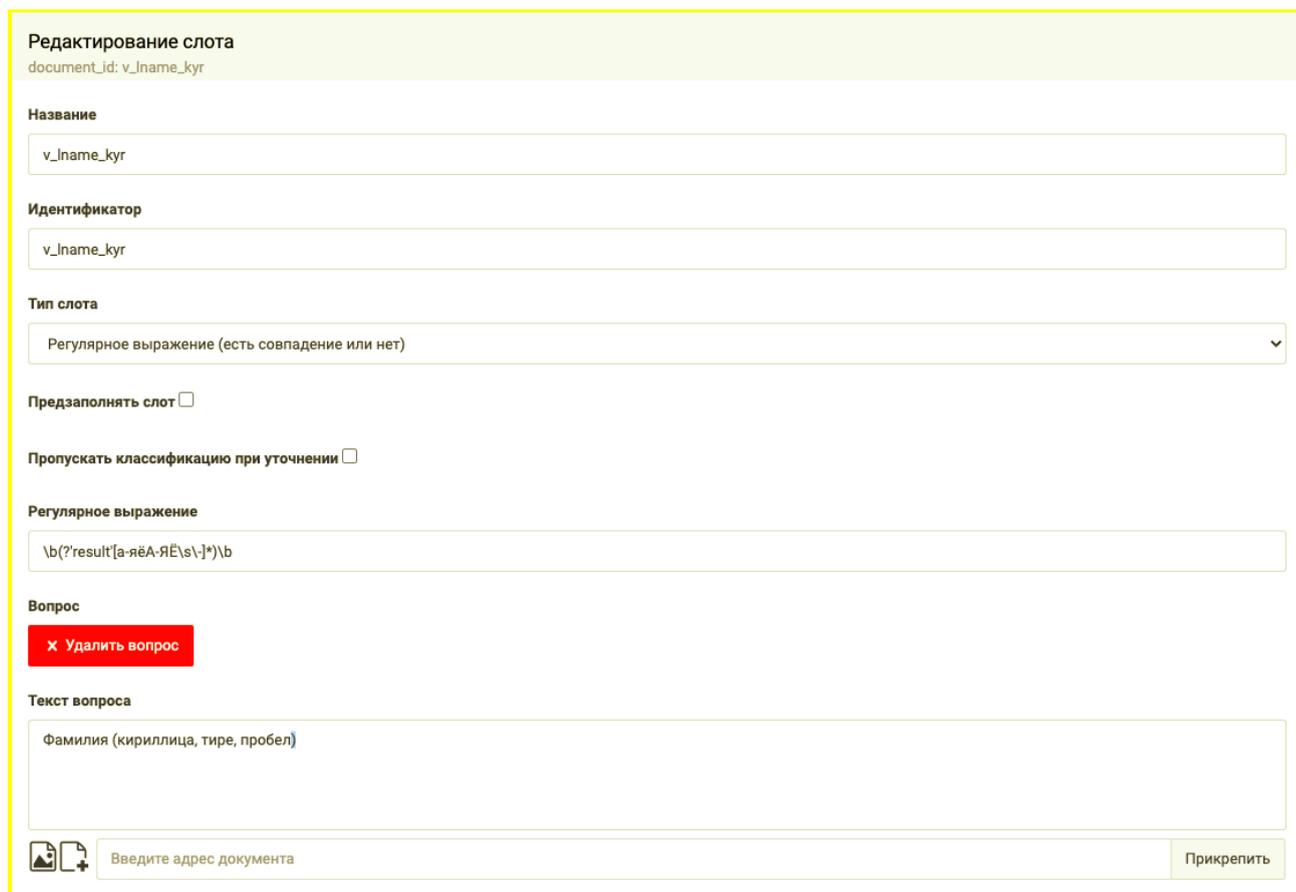
Рисунок 4.1

И заполняем их, например, как показано на рисунке 4.2.

Название и значение словаря могут быть введены на русском языке.

В «значение словаря» первый столбец – идентификатор, второй – текст, который будет отображаться на кнопке.

Рекомендация. Если значение словаря будет куда-то передаваться (передаваться будет идентификатор), то желательно делать его таким чтобы смысл значения был понятен. То есть удобней идентификатор и название кнопки делать одинаковыми (рис. 4.3), тогда, например, если передать на почту себе сообщение с идентификатором (передается Excel файлом) в тексте будет понятный по смыслу идентификатор.



Редактирование слота
document_id: v_iname_kyr

Название
v_iname_kyr

Идентификатор
v_iname_kyr

Тип слота
Регулярное выражение (есть совпадение или нет) ▾

Предзаполнять слот

Пропускать классификацию при уточнении

Регулярное выражение
`\b(?:result[a-яёА-ЯЁ\s\-\-]*)\b`

Вопрос
X Удалить вопрос

Текст вопроса
Фамилия (кириллица, тире, пробел)

 Введите адрес документа Прикрепить

Рисунок 4.2

Редактирование слота
id: c4ed1852-b754-4482-919e-5ac79972caa2

Название:

Идентификатор:

Тип слота:

Предопределить слот

Прогнозировать классификацию при уточнении

Показывать кнопку оценки

Не показывать кнопки словаря

Значения словаря

| | | | |
|------------|------------------|-------------|--------------------------|
| id: client | Юридическое лицо | Словарь (0) | <input type="checkbox"/> |
| id: client | Физическое лицо | Словарь (0) | <input type="checkbox"/> |

Вопрос

Текст вопроса

Вы являетесь юридическим или физическим лицом?

Приветствие

Второй вопрос

Вопрос для получения следующего значения

Отмена

Рисунок 4.3

После создания слотов надо создать анкеты.

Анкеты для сценариев создаются под название survey с номером (это уже не обязательное требование) (рис. 4.4).

Редактирование анкеты
id: c4ed1852-b754-4482-919e-5ac79972caa2

Общая информация | Настройки доступа

Название:

Поля анкеты

| | | | | | |
|-----|---------------|--------------------------------------|------|---|--------------------------|
| ↑ ↓ | Название поля | <input type="text" value="Текст"/> | Слот | <input type="text" value="тест: Слот 1 (текст)"/> | <input type="checkbox"/> |
| ↑ ↓ | Название поля | <input type="text" value="Словарь"/> | Слот | <input type="text" value="тест: Слот 2 (словарь)"/> | <input type="checkbox"/> |
| ↑ ↓ | Название поля | <input type="text" value="Число"/> | Слот | <input type="text" value="тест: Слот 3 (целое число)"/> | <input type="checkbox"/> |

Отмена

Рисунок 4.4

В Анкете заполняется «Название» - survey с номером, далее заполняется «Поле анкеты», это будет первый столбец в excel файле (ячейка A1), «Слот» выбирается из списка и будет является значение вносимым в первый столбец (A2). Следующее название поля будет соответственно столбец B1 в файле и его

значение В2. сгруппированы в анкету в том порядке, в котором в файле должны быть добавлены столбцы на лист-опросник (рис. 4.5).

| | A | B | C | D | E | F |
|----|---------------------------------------|-----------------------|----------------------|--------------|-----------------------------------|--|
| 1 | id диалекта | ФИО клиента | Идентификатор канала | Тип канала | id оператора | ФИО оператора |
| 2 | 39e93629-aa95-4aa6-9623-f4ac8033abb8d | Вика | webchat_alpha | webchat | DEFAULT_test1DEFAULT_vikki | Принят ботомПринят ботомПринят ботом |
| 3 | 0c59ae3d-e2d4-4f77-bdd2-b04f928a1eba | Вика Иванова | webchat_alpha | webchat | DEFAULT_vikkiDEFAULT_vikkiDEFAULT | Принят ботомПринят ботомПринят ботом |
| 4 | b6a78c69-2aeb-405e-95b2-8f6eac4a5a52 | Михаил Тест | webchat_alpha | webchat | DEFAULT_test1DEFAULT_vikki | Принят ботомПринят ботомПринят ботом |
| 5 | 147c88e9-c82f-4354-b37b-ec99dfafa17 | | alice_alpha | yandex_alisa | | Принят ботом |
| 6 | 6dd2a0f1-7adc-476a-b5a8-5e46167cdfa3 | Сергей Жандаров | webchat_alpha | webchat | | Принят ботом |
| 7 | dcf7d049-c2ed-4b33-9c75-bc0488778553 | Вика Иванова | webchat_alpha | webchat | DEFAULT_testDEFAULT_vikkiDEFAULT | Принят ботомПринят ботомПринят ботом |
| 8 | 58a44ed6-1926-4e6a-ad73-f2e04c54b294 | Вик | email_alpha | email | DEFAULT_vikki | Виктория Тест |
| 9 | 89570e47-caf2-43e5-951c-958703bf724a | Олег Агапов | webchat_alpha | webchat | DEFAULT_test | Принят ботомПринят ботомПринят ботом |
| 10 | 65d879db-f97b-4787-bf6e-4eefb55a044e | Denis Petukhov | telegram_alpha | telegram | | Принят ботом |
| 11 | fe636d5f-0fb7-423a-9c66-4365330be12a | | alice_alpha | yandex_alisa | | Принят ботом |
| 12 | 7b2e0500-e790-4979-b319-38e0c86e50e0 | Анна Иванова | webchat_alpha | webchat | DEFAULT_vikkiDEFAULT_vikkiDEFAULT | Принят ботомПринят ботомПринят ботом |
| 13 | 67f2f573-3f5b-4a0e-8e90-d42be41b6591 | Виктория Тестирование | telegram_alpha | telegram | DEFAULT_vikki | Принят ботомПринят ботомВиктория Тест |
| 14 | a5712a43-840e-4d5f-85f3-85aba54a6dc0 | Вика | webchat_alpha | webchat | DEFAULT_vikki | Принят ботомПринят ботомПринят ботом |
| 15 | 83370f2b-b5d4-4cb9-8fc8-bb7d59e0c513 | Вика | webchat_alpha | webchat | DEFAULT_vikki | Принят ботомПринят ботомПринят ботом |
| 16 | bd43f157-86bd-4b50-99ea-f0192a509a93 | Вика | webchat_alpha | webchat | DEFAULT_vikki | Принят ботомПринят ботомПринят ботом |
| 17 | 57b6d4f6-5d07-4521-8330-c62a10daf22a | Иван Иванов | webchat_alpha | webchat | DEFAULT_test1DEFAULT_vikkiDEFAULT | Принят ботомПринят ботомПринят ботом |
| 18 | 631eca0b-6f1c-4bae-b012-efa7354ec7b5 | Вика Тест | email_alpha | email | DEFAULT_vikki | Виктория Тест |
| 19 | 1d3d8393-0414-4ccc-b179-33448777178d | Вика Тест | webchat_alpha | webchat | DEFAULT_vikkiDEFAULT_vikki2 | Принят ботомПринят ботомПринят ботом |
| 20 | 1d0db16d-11a0-413f-ad02-dd9466cb60b3 | Виктория Тестирование | telegram_alpha | telegram | DEFAULT_vikkiDEFAULT_vikki | Принят ботомВиктория ТестВиктория Тест |
| 21 | 6b0e57eb-3918-4209-8a38-1203e66061bd | Вика Тест | webchat_alpha | webchat | DEFAULT_vikkiDEFAULT_vikkiDEFAULT | Принят ботомПринят ботомПринят ботом |
| 22 | 09456e35-05d3-4d96-802a-18c080e79ae1 | Вика Тест | webchat_alpha | webchat | DEFAULT_vikki | Принят ботомПринят ботомПринят ботом |
| 23 | a1445360-2235-4ba7-84c4-798a6ccce59a | | alice_alpha | yandex_alisa | | Принят ботом |
| 24 | f11d89d0-8e7d-4205-8b7d-6662e100fc74 | | alice_alpha | yandex_alisa | | Принят ботом |

Рисунок 4.5

Далее вся информация о названиях слотов и e-mail адрес куда будет отправляться информация передается программисту, который подключит агента, которого нельзя просто добавить. Программист в настройках добавит нового агента `surveys_agent(номер)` с его помощью передается содержимое слотов в виде файла Excel (рис. 4.6) (`surveys_agent` не поставляется в системе по умолчанию и является агентом, добавляемым по запросу).

Агент `surveys_agent1 (surveys_agent1)`

Идентификатор

Агент в случае неудачи текущего

Следующий агент

Доп. идентификатор агента

Возвращаемый агентом текст

Рисунок 4.6

В строку «Возвращаемый агентом текст» можно прописать строку `/switch_redirect_AIAssist_ID` либо `<имя_агента>`. То есть указать куда мы собираемся перейти, если сбор анкеты не является последним шагом (рис. 4.7).

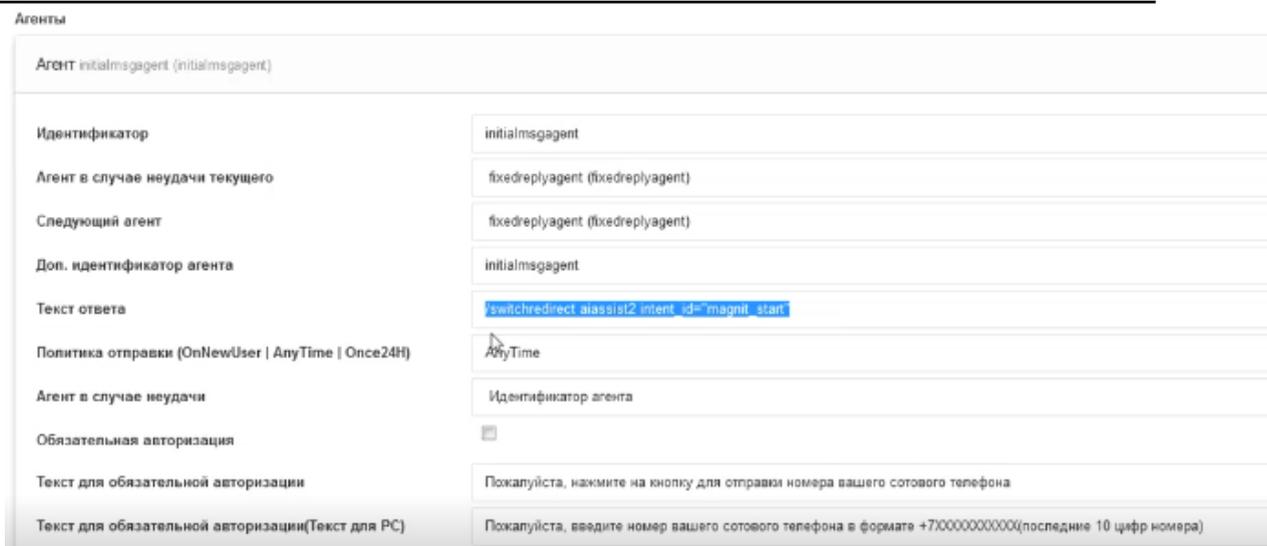


Рисунок 4.7

Это означает что пользователь получит сообщение и будет знать, что его информация получена.

После этого произошёл возврат к Боту.

Третьим шагом создаем статью Базы знаний (для созданного survey) с реакцией на проблему (рис. 4.8).

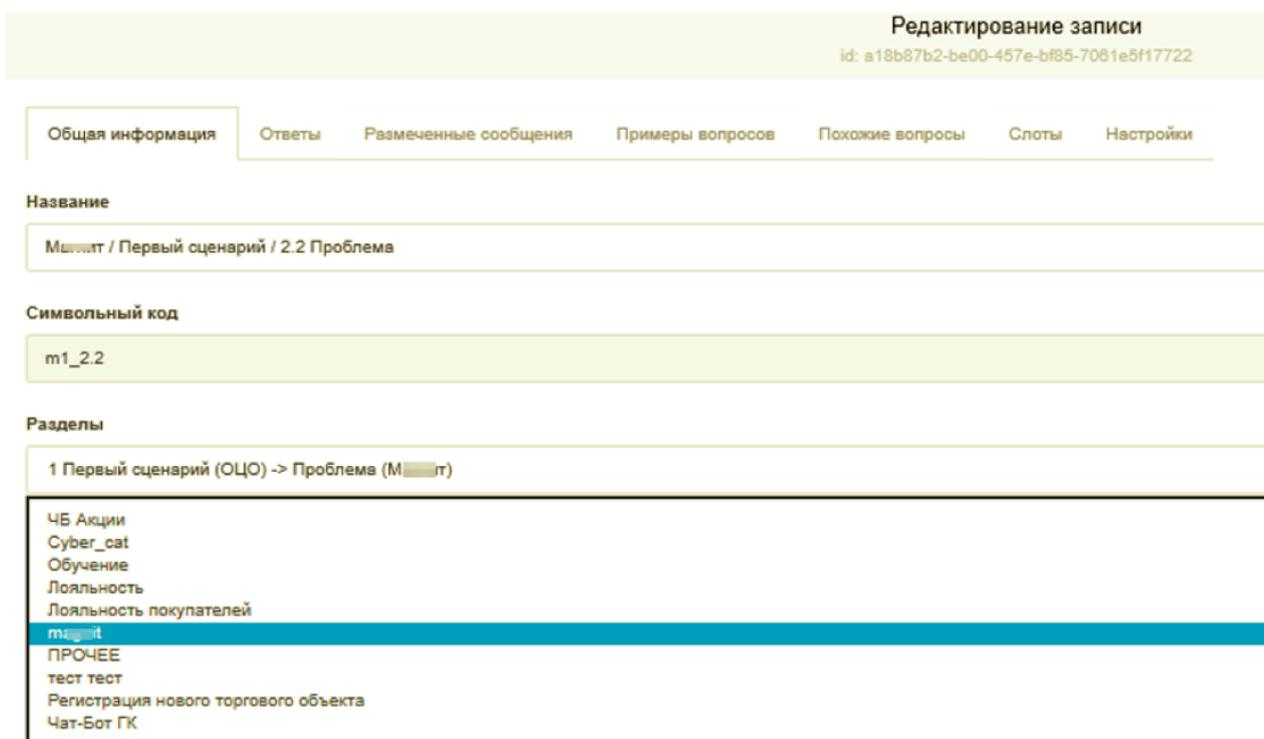


Рисунок 4.8

Заполняем «Название» и «Символьный код». Поле «Разделы» проставляется для разделения статей БЗ, здесь сохраняется тематика, к которой относится данная статья Базы знаний.

При наборе текста предполагаемого названия в строке появляется таблица вариантов тематик и подтематик. Например, если при введении текста «магнит» появится подтематика – Первый сценарий (ОЦО) (рис. 4.9).

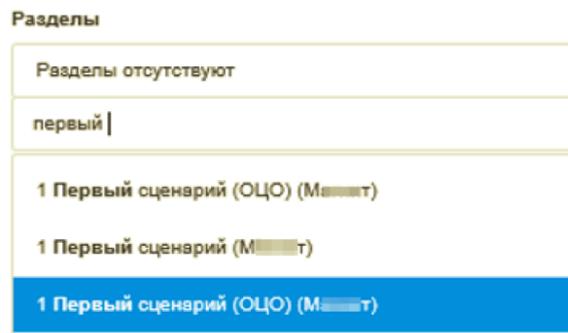


Рисунок 4.9

Переходим в «Ответы» и ответ с вариантами в зависимости от нажатой кнопки.

Вводим вопрос и кнопки к нему для дальнейшего ветвления ответа (рис. 4.10).

Ответы

ОТВЕТ Непременно разберемся! С чем связана ...

Тип текста ответа

Простой текст HTML

HTML-форматирование доступно только для информационных статей и ответов в канале Webchat

Файл ▾ Изменить ▾ Вид ▾ Вставить ▾ Формат ▾ Инструменты ▾ Таблица ▾ Помощь ▾

 Формат ▾
  Параграф ▾
  Шрифт ▾
  **B**
 *I*
 ~~S~~
 A ▾
 A ▾




















Непременно разберемся!

С чем связана Ваша проблема?

  Введите адрес документа

Кнопки:

Поиск записи в базе знаний

| | | |
|--------------|----------------------------------|-------------|
| m1_2.2.1.1.1 | Не пришла заработная плата/аванс | Синонимы(8) |
| m1_2.2.1.2.1 | Премия | Синонимы(2) |
| m1_2.2.1.3.1 | Отпуск | Синонимы(1) |
| m1_2.2.1.4.1 | Больничный лист | Синонимы(6) |
| m1_2.2.1.5.1 | Увольнение | Синонимы(2) |
| m1_2.2.1.6.1 | Налоговые вычеты | Синонимы(3) |
| m1_2.2.1.7.1 | Отпуск по уходу за ребенком | Синонимы(0) |
| m1_2 | Назад | Синонимы(0) |

Рисунок 4.10

Для создания новой кнопки требуется ввести идентификатор существующей (или будущей) статьи в БЗ и название, которое будет отображаться на самой кнопке (“Текст кнопки”).

Для удобства поиска записей в БЗ, на которые будут ссылаться кнопки, предусмотрена строка поиска. В нее достаточно ввести название записи (символьного кода). В случае успешного поиска будет выведена подсказка, на которую достаточно кликнуть мышкой, и поля идентификатор и текст кнопки заполнятся автоматически.

При выборе одной из кнопок пользователем ему будет выведена соответствующая запись либо произойдёт очередное ветвление, если оно присутствует в записи (рис. 4.11).

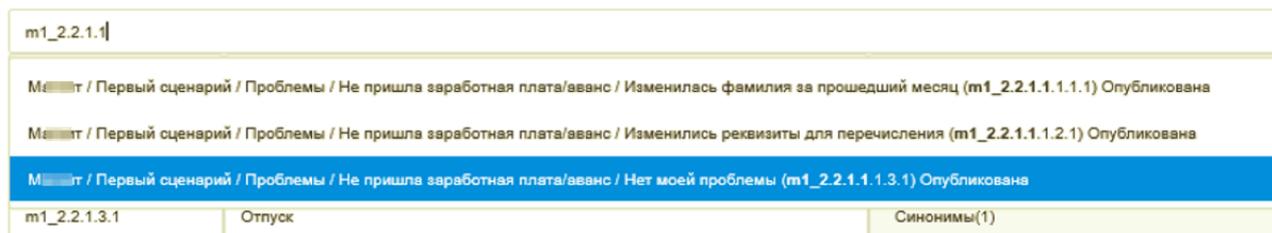


Рисунок 4.11

Приложение 1. Классификаторы

1.1 Обучение чат-бота (классификатор WMD)

Как правильно обучать чат-бота чтобы он реагировал на ключевые слова и понимал, что от него хотят? Например, чтобы для каждого из таких вопросов были разные статьи и бот не путался?

1. Оформить кредитную карту.
2. Кредитные карты.
3. Подобрать кредитную карту.

Так как же сделать так, чтобы чат-бот не реагировал на 100% точные формулировки из Примеров вопросов? Например, есть обученная статьи с примерами вопросов типа как "Как оформить кредитную карту?", но, если человек введет "Как я могу оформить кредитную карту?" - бот не находит нужного ответа, хотя 80% слов совпадает.

Ответ:

Главная причина, по которой бот путается в ответах - это дубликаты - те фразы, которые бот считает одинаковыми. Допустим, клиент задал вопрос. Wmd ищет ближайшую фразу, наиболее похожую на вопрос. Если ближайшей к вопросу фраза присутствует в двух разных интентах, то бот получит одинаковую вероятность того, что любой из этих интентов верный, и значит вернет любой из них.

Дубликаты для вопросов про кредитные карты показаны на рисунке П1.1.

| | questions | labels | lemmas |
|------|-------------------------------|------------------------------------|--------------------------------|
| 342 | кредитная карта с кешбек | natural_cards_creditcard_owner | кредитный_A карта_S |
| 1922 | своя кредитная карта | owner_credit_card2 | кредитный_A карта_S |
| 1923 | кредитная своя карта | owner_credit_card2 | кредитный_A карта_S |
| 2112 | Вопрос по кредитной карте. | natural_cards_creditcard | кредитный_A карта_S |
| 2114 | выбрать кредитную карту | natural_cards_creditcard | кредитный_A карта_S |
| 2139 | Кредитные карты | natural_cards_creditcard | кредитный_A карта_S |
| 2123 | Получить кредитную карту. | natural_cards_creditcard | получать_V кредитный_A карта_S |
| 2657 | Где получить кредитную карту? | natural_cards_creditcard_card-draw | получать_V кредитный_A карта_S |

Рисунок П1.1

Здесь lemmas - это то, как фразу видит бот. Заметьте, что это та же фраза, только слова в начальной форме, и некоторые выброшены. Можно выделить 2 ключевых этапа, на которых из фразы выпадают слова: лемматизация и фильтрация лемм, отсутствующих в эмбедингах.

1. Лемматизация. На этом этапе в предложении находятся слова и приводятся в начальную форму. Затем выбрасываются некоторые части речи, такие как предлоги и местоимения. После выбрасываются стоп-слова - слова, которые обычно не относятся к сути вопроса. Например, во 2 и 3 строчке выброшено местоимение "своя". В 4 и 5 -- стоп-слова "вопрос" и "выбрать".

2. Фильтрация лемм, отсутствующих в эмбедингах. Есть словарь эмбедингов. Если слово в нем отсутствует, то бот его не видит. В этом словаре содержится большинство русских слов. Плюс, туда также добавляются специфичные термины, индивидуальные для заказчика. Слово добавляется, если прошло первый этап, то есть для него удалось подобрать лемму, и оно не отсеялось, как незначущая часть речи и если оно содержится в датасете - в вопросах и ответах - не менее 3 раз. Например, по этому принципу было отброшено слово "кешбек" в 1-ой строчке. Оно определяется как лемма "кешбек_S". Его нет в базовых эмбедингах, и оно не добавляется в словарь, так как встречается в датасете всего 2 раза.

То есть, чтобы бот не путался, должно быть как можно меньше похожих примеров в разных интентах.

Что касается фразы "Как я могу оформить кредитную карту?". Бот видит ее как 4 слова: ['мочь_V', 'оформлять_V', 'кредитный_A', 'карта_S'].

WMD точного совпадения не находит, поэтому ищет примеры, в которых совпадает 3 слова из 4 и сравнивает расстояние между отличающимися четвертыми. Важность слов одинакова. Поэтому для WMD хорошие варианты ближайшей фразы не только "Как оформить кредитную карту?" (оформлять_V кредитный_A карта_S), но и фразы из других интентов:

- "Где я могу закрыть кредитную карту?" (мочь_V закрывать_V кредитный_A карта_S);

- "я могу получить кредитную карту." (мочь_V получать_V кредитный_A карта_S).

Решением здесь могло бы быть добавление леммы "мочь_V" в стоп-слова, чтобы она выбрасывалась на этапе лемматизации. К сожалению, это нельзя сейчас сделать через графический интерфейс, но мы можем сделать это вручную в коде.

GRU же даже не рассматривает интен `natural_cards_creditcard_card-draw`, которому принадлежит фраза, так как он слишком мал для него. GRU классифицирует только среди интен `тов`, у которых есть минимум 20 примеров.

Возвращаясь к вопросу:

Есть 3 статьи.

1. Оформления кредитной карты
2. О кредитных картах
3. Подбор кредитной карты

Какие примеры вопросов должны быть в каждой из них, чтобы их отличать? (по 1-2 примера достаточно)

Ну что тут посоветовать: делать примеров больше и разнообразнее. Главное, чтобы леммы, которые остаются после препроцессинга, для разных интен `тов` не пересекались.

Леммы = Все слова - стоп-слова - слова_без_эмбедингов

То есть если убрать стоп-слова и слова_без_эмбедингов из предложения, то предложения из разных интен `тов` все еще должны отличаться.

Со словами без эмбедингов справиться легко. Если добавляете пример с необычным, редким словом, которого может не быть в словаре, то нужно создать минимум 3 примера с его использованием. Также следует иметь в виду, что у таких слов могут также возникнуть проблемы с приведением в начальную форму, то есть хорошо бы добавить по 3 примера на каждую форму слова, которую хотите использовать.

Со стоп-словами главное обращать внимание, чтобы стоп-слова не были ключевыми в примере. Если в одном интен `те` "кредитная карта", а в другом "вопрос

о кредитной карте", то будут пересекаться, так как здесь стоп-слово "вопрос" - ключевое - единственное, которым отличаются.

Или наоборот, бывают проблемы, когда какое-то слово, которое можно было бы выбросить как стоп-слово не выбрасывается. Как например "могу". Тут, мне кажется, лучше всего, если добавляется пример с таким незначащим словом в какой-то интент, то нужно добавить аналогичный пример с этим словом в каждый похожий интент.

В итоге примеры для 3 статей:

- Оформление кредитной карты
- Оформление кредитки
- Оформить кредитную карту
- Оформить кредитку
- Как я могу оформить кредитную карту
- Как я могу оформить кредитку

- О кредитных картах
- Кредитная карта
- кредитка

- Подбор кредитной карты
- Подбор кредитки
- Подобрать кредитную карту
- Подобрать кредитку
- Как я могу подобрать кредитную карту
- Как я могу подобрать кредитку

Внимание! Сейчас "могу" и "хочу" добавлены в стоп-слова.

1.2 Примеры вопросов не на языке классификатора. Английские слова в примерах для обучения

Чтобы какое-то специфическое слово попало в обучение (слово на иностранном языке, какое-то сокращение, аббревиатура), необходимо включить в обучающие фразы не менее трех примеров с данным словом.

Английские слова, когда язык классификатора русский, регистрозависимы, то есть *ivi* и *IVI* - это разные слова, то есть каждое из них должно встречаться 3+ раз в примерах вопросов.

Переобучение эмбеддингов нужно, если стали использовать новое необычное слово или слово на английском, т.е. при переобучении требуется проставить соответствующую галку переобучить эмбеддинги.

1.3 Обучение классификаторов. Эмбеддинги. Переобучение эмбеддингов

Эмбеддинги - это смыслы слов, записанные числами (рис. П1.2).

4-х мерный Embedding

| | | | | |
|-----------------|-----|------|------|-----|
| кот => | 1.2 | -0.1 | 4.3 | 3.2 |
| круто => | 0.4 | 2.5 | -0.9 | 0.5 |
| вау => | 2.1 | 0.3 | 0.1 | 0.4 |
| ... | | | | ... |

Рисунок П1.2

В базовых эмбеддингах есть большинство обычных слов, которыми люди пользуются чаще всего. Когда добавляется классификатор, эмбеддинги дообучаются на данных заказчика: дополняются его словами, которых нет в базовых, смыслы немного корректируются, чтобы соответствовать контексту заказчика.

Переобучение эмбеддингов нужно, если стали использовать новое необычное слово или слово на английском. Должно быть минимум 3 примера с этим словом, чтобы для него создавался эмбеддинг. Также полезно переобучить эмбеддинги, если появилось много новых примеров - это улучшит смыслы.

В остальных случаях переобучение эмбеддингов необязательно. Если оно все же будет проставлено, то обучение будет длиться чуть дольше и будет чуть

больше зависеть от случайности, то есть качество классификации может стать чуть лучше или хуже.

Больше всего от эмбедингов зависит классификатор WMD. Чуть меньше - GRU (CNN) и KNN. Совсем не использует эмбединги в данный момент logreg.

1.4 Описание классификатора KNN

Что такое KNN? KNN — классификатор K Nearest Neighbors (K ближайших соседей).

Чтобы присвоить класс сообщению пользователя, KNN:

- Находит K наиболее похожих примеров в обучающей выборке;
- Смотрит, примеры какого класса встречаются в них чаще всего.

Это и будет результатом классификации.

Score результата приблизительно равен доле самого популярного класса в K ближайших.

Сейчас $K = 4$. Что будет, если среди 4 ближайших соседей 2 будут принадлежать к одному к классу и 2 к другому? Оба класса получат score примерно равный 0.5. Большой score будет у того класса, примеры которого ближе к сообщению пользователя.

При настройке порогов имеет смысл перебирать пороги близкие к 0.5 и 0.75 (рис. П1.3).

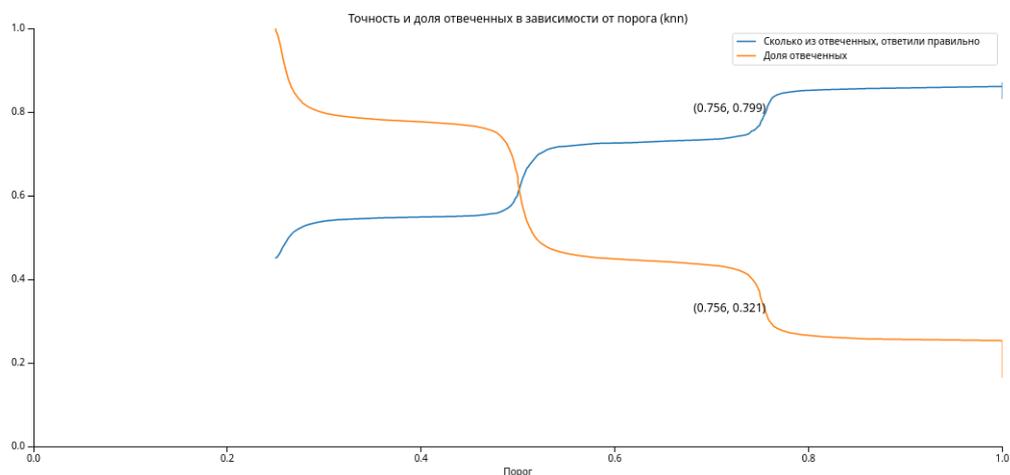


Рисунок П1.3

Свойства KNN

Благодаря тому, что KNN ищет ближайших соседей, он лучше работает с маленькими интентами, чем GRU или LogReg, примерно как WMD.

Также он лучше этих классификаторов запоминает обучающую выборку. То есть на вопросы, которые добавили в примеры, он должен отвечать в 98%+ случаев.

По скорости ответа KNN выигрывает у WMD. WMD, кстати, работает схожим образом - ищет самое близкое предложение в обучающей выборке. Но KNN быстрее считает расстояние между примерами.

Результаты KNN сложнее интерпретировать, чем результаты WMD. Для обоих можно сказать, какие примеры оказались ближайшими, но для KNN человеку непонятно, почему именно эти примеры — ближайšie, тогда как для WMD обычно ясно.

Эмбеддер или как измерить близость?

Как понять, насколько один текст похож на другой? Преобразуем текст в числовой вектор, который будет "содержать смысл" текста. Это значит, что схожие по смыслу текста, будут иметь близкие числовые вектора. Тогда для того, чтобы сравнить похожесть текстов, можно будет сравнить похожесть числовых векторов.

Возможно вы знаете, что для работы большинства классификаторов у нас используются эмбеддинги — численные представления слов. Для работы KNN тоже нужны эмбеддинги. Только не эмбеддинги слов, а эмбеддинги целых предложений. Это как раз те числовые вектора, о которых я писал в абзаце выше.

Модель, которая делает эмбеддинг предложения, называется эмбеддер. Вообще, эмбеддером может быть любая модель, у которой на определенной стадии появляется сравнимое численное представление предложения.

Наш основной эмбеддер основан на GRU. GRU — это другой классификатор, нейронная сеть. В качестве эмбеддинга предложений GRU эмбеддер возвращает слой перед выходными нейронами, так как он содержит всю информацию, которую классификатор смог извлечь из предложения. ***Поэтому для работы KNN необходим классификатор GRU на том же сервере.***

Чтобы связать KNN и GRU нужно:

-
- имя GRU классификатора должно быть flat_cpp;
 - директории датасета должны совпадать;
 - Builder-ы для препроцессинга должны совпадать;
 - заказчик должен быть один и тот же.

Помимо GRU эмбеддера, в разработке есть также эмбеддер, основанный на BERT — более современной нейронной сети. В будущем, если ресурсы сервера позволяют, его также можно будет использовать.

Обучение KNN

Запустить обучение KNN, как и любого другого классификатора можно двумя способами: добавив новый классификатор и нажав кнопку "Переобучить" у существующего.

Создание классификатора

Создать раньше GRU или KNN — неважно, но необходимо дождаться завершения обучения первого классификатора, прежде чем создавать второй. Главное — правильно их связать (см. выше).

Если создать KNN до GRU, то GRU будет создано автоматически, но скрытно. Это значит, что оно не появится в UI, но будет использоваться для KNN. Но не создавать GRU не рекомендуется, так как без него нельзя будет переобучить эмбеддинги слов и предложений, и тогда со временем KNN станет работать хуже.

Галочку "Переобучить эмбеддинги" здесь лучше нигде не ставить. Для первого классификатора они обучатся автоматически. Для второго возьмутся из первого.

Переобучение

KNN переобучается быстро, так что его можно переобучать, когда угодно. Можно по несколько раз на день. Галочку "Переобучить эмбеддинги" не ставить.

GRU переобучается долго, так что его обычно переобучают по расписанию. GRU лучше переобучать с галочкой "Переобучить эмбеддинги", чтобы он подцепил все новые слова. После того, как GRU закончит переобучаться, нужно переобучить KNN.

Также, если сбросилось значение Builder для препроцессинга, не забудьте его указать.

Пример правильных настроек KNN и GRU

KNN:

| Основные параметры | |
|--|---|
| Заказчик | <input type="text" value="mycustomer"/> |
| Тип классификатора | <input type="text" value="KNN"/> |
| Язык классификатора | <input type="text" value="Русский"/> |
| Хост сервера классификации | <input type="text" value="http://qaclassifiers-qaclassifiers:3000"/> |
| Директория датасета | <input type="text" value="/data/flat_cnn"/> |
| Builder для препроцессинга | <input type="text" value="preprocessor_builders.lower"/> |
| Переобучить эмбединги | <input type="checkbox"/> |
| Какие данные использовать | |
| Записи базы знаний (интенты) | <input checked="" type="checkbox"/> |
| Адрес elasticsearch | <input type="text" value="http://elasticsearch-elasticsearch-es-http.elastic:9200/mycustomer_i"/> |
| Сообщения (размеченные операторами) | <input checked="" type="checkbox"/> |
| Адрес elasticsearch | <input type="text" value="http://elasticsearch-elasticsearch-es-http.elastic:9200/mycustomer_i"/> |
| Фильтры данных для обучения | |
| Фильтры отсутствуют | |
| <input type="button" value="Добавить фильтр"/> | |

GRU:

Основные параметры

| | |
|-------------------------------------|---|
| Заказчик | <input type="text" value="mycustomer"/> |
| Тип классификатора | <input type="text" value="CNN"/> |
| Язык классификатора | <input type="text" value="Русский"/> |
| Хост сервера классификации | <input type="text" value="http://qaclassifiers-qaclassifiers:3000"/> |
| Директория датасета | <input type="text" value="/data/flat_cnn"/> |
| Builder для препроцессинга | <input type="text" value="preprocessor_builders.lower"/> |
| Переобучить эмбединги | <input type="checkbox"/> |
| Какие данные использовать | |
| Записи базы знаний (интенты) | <input checked="" type="checkbox"/> |
| Адрес elasticsearch | <input type="text" value="http://elasticsearch-elasticsearch-es-http.elastic:9200/mycustomer_i"/> |
| Сообщения (размеченные операторами) | <input checked="" type="checkbox"/> |
| Адрес elasticsearch | <input type="text" value="http://elasticsearch-elasticsearch-es-http.elastic:9200/mycustomer_i"/> |
| Фильтры данных для обучения | |
| Фильтры отсутствуют | |
| | <input type="button" value="Добавить фильтр"/> |

Проблемы хостинга

KNN не создается

Вероятно баг. Обращайтесь к разработчикам с логами qaclassifiers. (ML/AIAssist team, Nikita Furin)

KNN неправильно классифицирует простые вопросы

Возможно, напутали с галочками "Переобучить эмбеддинги". Переобучите GRU с эмбедингами, а затем, когда оно полностью обучится, переобучите KNN без эмбедингов.

1.5 Время и скорость обучения классификаторов

CNN - очень долго обучается до суток (ставить на ночь).

KNN - работает на модели CNN поэтому без нее не запускать (обучается до 30 минут).

LOGREG - обучается до 30 минут работает на messages и intents.

WMD - обучается до 10 минут работает в основном на intent (но иногда и messages).

WMD можно (и лучше) обучать и на intent и на messages. Но нужно контролировать чтобы было не слишком много примеров для обучения (до 500, ну может до 1000) - иначе при выдаче результатов будет тормозить и есть CPU.

Для остальных классификаторов количество примеров влияет только на скорость обучения и слабо влияет на скорость выдачи (или вообще не влияет).

1.6 Список стоп-слов чат-бота

Для русского игнорируются:

['и', 'в', 'во', 'что', 'он', 'на', 'я', 'с', 'со', 'а', 'то', 'все', 'она', 'так', 'его', 'но', 'ты', 'к', 'у', 'же', 'вы', 'за', 'бы', 'по', 'только', 'ее', 'мне', 'было', 'вот', 'от', 'меня', 'еще', 'о', 'из', 'ему', 'теперь', 'когда', 'даже', 'ну', 'вдруг', 'ли', 'если', 'уже', 'или', 'быть', 'был', 'него', 'до', 'вас', 'нибудь', 'опять', 'уж', 'вам', 'ведь', 'там', 'потом', 'себя', 'ничего', 'ей', 'может', 'они', 'тут', 'где', 'есть', 'надо', 'ней', 'для', 'мы', 'тебя', 'их', 'чем', 'была', 'чтоб', 'будто', 'чего', 'раз', 'тоже', 'себе', 'под', 'будет', 'ж', 'тогда', 'этот', 'того', 'потому', 'этого', 'какой', 'совсем', 'ним', 'здесь', 'этом', 'один', 'почти', 'мой', 'тем', 'чтобы', 'нее', 'были', 'куда', 'зачем', 'всех', 'никогда', 'при', 'наконец', 'два', 'об', 'другой', 'хоть', 'после', 'над', 'больше', 'тот', 'через', 'эти', 'нас', 'про', 'всего', 'них', 'какая', 'много', 'разве', 'три', 'эту', 'моя', 'впрочем', 'хорошо', 'свою', 'этой', 'перед', 'иногда', 'лучше', 'чуть', 'том', 'такой', 'им', 'более', 'всегда', 'конечно', 'всю', 'между', 'подсказывать', 'помогать', 'выбирать', 'уточнять', 'вопрос', 'возможно', 'нужный', 'хотеть', 'мочь']

Также игнорируются все союзы, частицы, предлоги и местоимения **за исключением**

["не", "нет", "ни", "да", "можно", "нельзя", "сейчас", "здрaсте", "сам", "без", "кто", "как"]

Для английского игнорируются:

['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above', 'below', 'to', 'from', 'in', 'on', 'again', 'further', 'then',

'once', 'here', 'there', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'other', 'some', 'such', 'only', 'so', 'than', 'very', 's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan', "shan't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn', "wouldn't"]

Также вспомогательные части речи (coordinating conjunction, subordinating conjunction, adposition, determiner, symbol), **за исключением:**

['no', 'nor', 'not', 'yes', 'up', 'down', 'off',
'now', 'how', 'out', 'over', 'own', 'same', 'too',
'under', 'until', 'what', 'when', 'where', 'which',
'while', 'who', 'whom', 'why']

Приложение 2. Добавление ответов для различных каналов

2.1 Добавление ответов для различных каналов в рамках одной статьи

Разделение ответов по каналам возможно настроить в том случае, если для каждого из ответов в рамках одной статьи будет добавлен слот “Каналы”.

Чтобы иметь возможность воспользоваться слотом “Каналы”, его необходимо создать в меню “База знаний” -> “Слоты”. Слот создается **один** раз и далее используется в статьях при необходимости разделения ответов в рамках одной статьи по каналам.

2.2 Создание слота “Каналы”

1. Перейти в меню **База знаний** на вкладку **Слоты**.
2. Нажать кнопку “Добавить” и заполнить поля открывшейся формы:
 - поле **Название**: Каналы
 - поле **Идентификатор**: channel_id
 - поле **Тип слота**: Словарь (выбрать из выпадающего списка)
 - Нажать на панель **Значения словаря**, далее нажать на кнопку

Добавить значение и заполнить поля **Идентификатор** и **Название** для каждого из каналов, которые применяются в системе (в случае, если каналов более одного, нажатие на кнопку **Добавить значение** повторить).

Примерный перечень идентификаторов каналов:

- o webchat_(постфикс) (WebChat)
- o viber_(постфикс) (Viber)
- o email_(постфикс) (mmclient@go.customer.ru)
- o email_respform (Email Respform)
- o mbeta (beta – постфикс заказчика)
- o ibeta

Важно! Вы можете изменять названия каналов, которые будут отображаться в выпадающих списках при последующем использовании слота Каналы, но **идентификатор слота** и **идентификаторы значения** словаря **изменению не подлежат**, они **обязательно** должны быть занесены в слот-словарь Каналы **без изменений**.

- после заполнения всех полей нажать кнопку Создать внизу формы

Название

Каналы

Идентификатор

channel_id

Тип слота

Словарь

Значения словаря

Значения словаря

| | | |
|---------------|---------------------|---|
| webchat_beta | WebChat |  |
| viber_beta | Viber |  |
| email_beta | mmclient@go.beta.ru |  |
| mbeta | mbeta |  |
| ibeta | ibeta |  |
| email_resform | Email Respform |  |

 Добавить значение

2.3 Создание нескольких ответов для нескольких каналов в рамках одной статьи

1. Зайти в пункт меню **База знаний**, вкладка **Записи базы знаний**
2. Открыть для редактирования статью, в которую следует добавить ответ для другого канала или создать новую статью, если она еще не создана в Базе знаний
3. Перейти на вкладку **Ответы**

Редактирование записи
id: 16fbd96d-342d-4314-ae4e-f48d0b1fa65f

Общая информация **Ответы** Размеченные сообщения Примеры вопросов Похожие вопросы Слоты Настройки

Ответы

Ответ 

Тип текста ответа

Простой текст HTML Markdown

Добавить шаблон

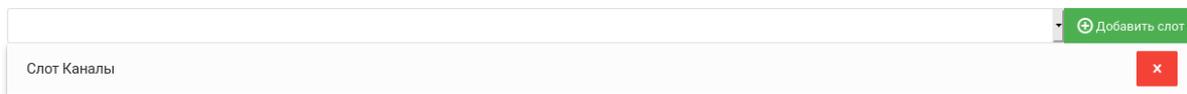
Укорачивание ссылок

Введите адрес ссылки

4. Создайте ответ, который должен уходить в один из каналов, например, в канал **WebChat** (кнопка **Добавить ответ**). Если ответ уже создан в рамках статьи, то ответ создавать не требуется, просто пропустите этот этап)

5. Добавьте слот Каналы под ответом для веб-чата - выберите из выпадающего списка **Значения слотов** пункт **Каналы** и нажмите **Добавить слот**.

Значения слотов:



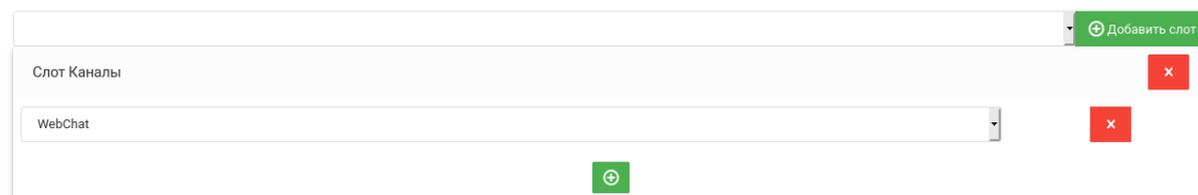
6. Раскройте панель **Слот Каналы** и нажмите на зеленую кнопку **+**

Значения слотов:



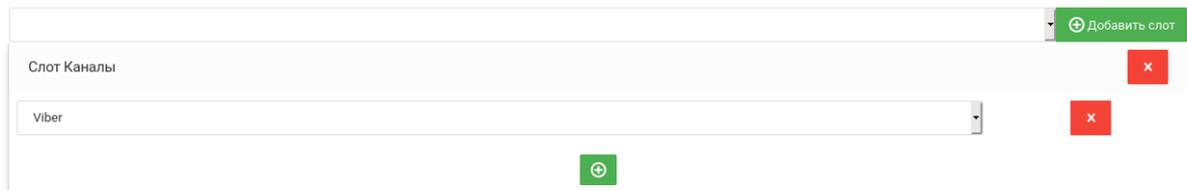
7. В появившемся выпадающем списке выберите канал **WebChat**

Значения слотов:



8. Для добавления ответа для другого канала, например, Viber, повторите п.п. 6-9, выбрав из выпадающего списка значение **Viber**:

Значения слотов:



9. В результате выполненных действий в статье будет создано два ответа, под каждым из которых будет установлен слот. Под первым ответом, для канала веб-чат, значение слота будет установлено как **WebChat**, а под вторым ответом, для канала вайбер, значение слота будет установлено как **Viber**.

10. Сохраните изменения в статье - нажмите **Создать черновик**. Далее следует отправить статью **на проверку** и **опубликовать**.

Когда система выберет в качестве ответа на вопрос клиента статью, в которой **есть разделение** ответов на каналы, если вопрос клиента **пришел из вебчата**, ему будет отправлен первый ответ (под которым в качестве значения слота **Каналы** выбрано **WebChat**), а если вопрос клиента придет из канала **viber**, то клиенту уйдет ответ, под которым установлен слот **Каналы** в значение **Viber**.

Приложение 3. Пример создания сценариев

На первом шаге на вкладке “Слоты” главного меню “База знаний” должны быть заведены слоты, которые потребуется для проведения опроса:

Слоты 

| Поиск | второ | | | |
|---|-------|----------|--|--------|
| Название | | Тип | | Кол-во |
| М...т / Второй сценарий / Файл да/нет | | Словарь | | 2 |
| М...т / Второй сценарий / Описание затрат | | Текст | | |
| М...т / Второй сценарий / Указание конечного подразделения | | Текст | | |
| М...т / Второй сценарий / Указание населенного пункта, департамента | | Текст | | |
| М...т / Второй сценарий / Указание почты и номера телефона | | Текст | | |
| М...т / Второй сценарий / Прикрепление файла | | Вложение | | |
| М...т / Второй сценарий / Указание темы идеи | | Текст | | |
| М...т / Второй сценарий / Описание сути идеи | | Текст | | |
| М...т / Второй сценарий / Описание проблемы | | Текст | | |
| М...т / Второй сценарий / Описание эффекта | | Текст | | |
| М...т / Второй сценарий / ФИО | | Текст | | |
| М...т / Второй сценарий / Выбор подразделения | | Словарь | | 11 |
| М...т / Второй сценарий / Выбор направления идеи | | Словарь | | 19 |

Анкета создается в меню “База знаний” на вкладке “Анкеты”. Слоты, содержимое которых будет передаваться с помощью агента survey4_agent в виде файла Excel, были сгруппированы в анкету в том порядке, в котором в файле должны быть добавлены столбцы на лист-опросник:

Редактирование анкеты
id: 99dbd83d-e00d-43e0-9aa5-3c872246bcb8

Общая информация
Настройки доступа

Название

survey4

Поля анкеты

| | | | | | |
|---|---------------|------------------------|------|--|---|
| ↓ | Название поля | ФИО | Слот | М...т / Второй сценарий / ФИО | 🗑 |
| ↓ | Название поля | Выбор направления идеи | Слот | М...т / Второй сценарий / Выбор направления идеи | 🗑 |
| ↓ | Название поля | Описание затрат | Слот | М...т / Второй сценарий / Описание затрат | 🗑 |
| ↓ | Название поля | Описание сути идеи | Слот | М...т / Второй сценарий / Описание сути идеи | 🗑 |
| ↓ | Название поля | Описание проблемы | Слот | М...т / Второй сценарий / Описание проблемы | 🗑 |
| ↓ | Название поля | Указание темы идеи | Слот | М...т / Второй сценарий / Указание темы идеи | 🗑 |
| ↓ | Название поля | Описание эффекта | Слот | М...т / Второй сценарий / Описание эффекта | 🗑 |

| | | | | | |
|---|---------------|---|------|--|---|
| ↓ | Название поля | Выбор подразделения | Слот | Меню / Второй сценарий / Выбор подразделения | ✕ |
| ↓ | Название поля | Указание конечного подразделения | Слот | Меню / Второй сценарий / Указание конечного подразделения | ✕ |
| ↓ | Название поля | Указание населенного пункта, департамента | Слот | Меню / Второй сценарий / Указание населенного пункта, департамента | ✕ |
| ↓ | Название поля | Указание почты и номера телефона | Слот | Меню / Второй сценарий / Указание почты и номера телефона | ✕ |
| ↓ | Название поля | Прикрепление файла | Слот | Меню / Второй сценарий / Прикрепление файла | ✕ |

[+ Добавить поле](#)

[Отмена](#) [Сохранить](#)

В качестве приветствия создаётся статья “М__т / Второй сценарий / начало” (символьный код m2_1). В ней содержится два ответа, один из которых будет отправлен в канал вебчат (для проверки этого условия под ответом добавлен слот “Каналы” с выбором webchat_aiassist, на вкладке “Слоты” внутри статьи добавлен слот “Каналы”), второй ответ уйдет в любой другой канал, кроме веб-чата, т.к. под ним не содержится условие. К каждому ответу добавлены кнопки, которые ведут на статьи в БЗ.

Одна из кнопок ведет на статью с описанием конкурса “М__т / Второй сценарий / Прочитать описание конкурса” (символьный код m2_1.1), вторая - на статью “М__т / Второй сценарий / Подать идею” (символьный код m2_1.2).

В статье про описание конкурса содержится несколько ответов. Это сделано для того, чтобы достаточно длинный текст не выходил одним полотном и лучше воспринимался людьми. В последнем ответе добавлены две кнопки - одна из них позволяет перейти к процессу подачи новой идеи (ссылка на статью “Подать идею”), а вторая ведет на статью “Подать идею позднее”, в которой выводится текст, призывающий сотрудников поделиться своими инициативами, когда они будут к этому готовы.

В статье “Подать идею” на вкладке “Слоты” добавлены заранее подготовленные слоты в том порядке, в котором информация будет запрашиваться у клиента:

| Общая информация | Ответы | Размеченные сообщения | Примеры вопросов | Похожие вопросы | Слоты | Настройки |
|------------------|---|-----------------------|------------------|-----------------|-------|-----------|
| Слоты | | | | | | |
| ↓ | Меню / Второй сценарий / ФИО | | | | | |
| ↓ | Меню / Второй сценарий / Выбор направления идеи | | | | | |
| ↓ | Меню / Второй сценарий / Указание темы идеи | | | | | |
| ↓ | Меню / Второй сценарий / Описание сути идеи | | | | | |
| ↓ | Меню / Второй сценарий / Описание проблемы | | | | | |
| ↓ | Меню / Второй сценарий / Описание эффекта | | | | | |
| ↓ | Меню / Второй сценарий / Файл да/нет | | | | | |
| ↓ | Меню / Второй сценарий / Прикрепление файла | | | | | |

На вкладке “Ответы” созданы три ответа. Два из них будут отправлены в канал клиента, если клиент ответит на все вопросы из слотов, в том числе на вопрос из слота “Файл да/нет”: “Хотите прикрепить примеры реализованной идеи в других сетях?” будет выбран ответ “Да” и отправлен файл-вложение. Один из этих ответов - сообщение, что “Файлы успешно загружены!”, а второй - переход на статью в БЗ с продолжением заполнения сценария с помощью команды (REDIRECT_TO_INTENT m2_z). Третий ответ будет выдан в случае, если клиент не имел намерения загружать вложение и на вопрос “Хотите прикрепить примеры реализованной идеи в других сетях?” дал отрицательный ответ - это ответ с командой-переходом на другую статью в БЗ (REDIRECT_TO_INTENT m2_z).

Статья “М__т / Второй сценарий / Описание затрат” (символьный код m2_z) содержит запрос информации по предстоящим затратам (с помощью слота “Описание затрат”). Если информация будет введена, то в рамках сценария будет выведена фраза с просьбой рассказать о сотруднике контактную информацию. При этом, для персонализации обращения, используется команда для вывода в тексте информации, содержащейся в слоте “ФИО” (символьный код m_name): __SLOT_VALUE_m_name. В качестве второго ответа в статье используется переход на статью “М__т / Второй сценарий / Подать идею 2” с помощью команды REDIRECT_TO_INTENT m2_1.2.test.

В статье “Подать идею 2” содержатся слоты, информация из которых должна быть запрошена у клиента (подразделение, конечное подразделение, населенный пункт и департамент, email и номер телефона), и слоты, которые ранее уже должны были быть запрошены на предыдущих этапах опроса:

| | |
|---|--|
| ↓ | М__т / Второй сценарий / Описание затрат |
| ↓ | М__т / Второй сценарий / Выбор подразделения |
| ↓ | М__т / Второй сценарий / Указание конечного подразделения |
| ↓ | М__т / Второй сценарий / Указание населенного пункта, департамента |
| ↓ | М__т / Второй сценарий / Указание почты и номера телефона |
| ↓ | М__т / Второй сценарий / ФИО |
| ↓ | М__т / Второй сценарий / Выбор направления идеи |
| ↓ | М__т / Второй сценарий / Описание сути идеи |
| ↓ | М__т / Второй сценарий / Описание проблемы |
| ↓ | М__т / Второй сценарий / Указание темы идеи |
| ↓ | М__т / Второй сценарий / Описание эффекта |
| ↓ | М__т / Второй сценарий / Файл да/нет |
| ↓ | М__т / Второй сценарий / Прикрепление файла |

Если клиент ответит на все вопросы (т.е. все слоты будут заполнены в ходе прохождения сценария), а факт заполнения мы проверяем выставлением условия `__FILLED__` под каждым из слотов, то в канал клиента будет отправлена фраза-благодарность за желание сделать компанию и мир вокруг лучше, в качестве второго ответа присутствует команда с переходом на статью в БЗ, в которой происходит вызов агента для передачи `xlsx`-файла на указанный в настройках агента `e-mail`-адрес. Переход на статью с помощью команды в рамках ответа выглядит так:

```
REDIRECT_TO_INTENT m2_1.2.11
```

В статье присутствует 4 ответа, по два на каждый вариант набора слотов. Варианта два, т.к. в одном случае клиент не посылает файл с вложением и отвечает “нет” на вопрос из слота “Файл да/нет”: “Хотите прикрепить примеры реализованной идеи в других сетях?”, а в другом случае отвечает “да” и появляется в наборе проверка заполнения слота типа “Вложение”.

В статье “М__т / Идея / Вызов агента” (символьный код `m2_1.2.11`), как было указано ранее, происходит вызов агента `surveys_agent4` для передачи ранее созданной анкеты `survey4` (данные в которую передаются из заполненных в ходе опроса слотов). Вызов агента происходит с помощью команды `/switchredirect <имя_агента>`:

```
/switchredirect surveys_agent4.
```

Также как идею можно подать много раз, а слот используется один и тот же необходимо создать запись с стиранием данного слота либо группы слотов.

База знаний

Поиск по всем разделам

Поиск записей

de|

Записи с 1 до 12 из 12 записей (отфильтровано из 4,699 записей)

| Название |
|--|
| Возврат анкеты |
| Очистка слота Выбрать год |
| Очистка слота Выбор месяца2019 |
| Очистка слота Выбор месяца2020 |
| Очистка слота месяц для Главное меню |
| Очистка слот Акции ГК |
| Очистка слота Выбрать гр занятие |
| Очистка слота выбора гр занятий Главное меню |
| portal_start_brdel |
| portal_start_br |
| Меню для сценариев |
| Меню / Идея / Удаление заполненных слотов |

Для этого создаётся запись.

Редактирование записи
id: 931f50dc-a4c5-43cf-b648-3ce2ae9b229c

Общая информация | Ответы | Размеченные сообщения | Примеры вопросов | Похожие вопросы | Слоты | Настройки

Название
Меню / Идея / Удаление заполненных слотов

Символьный код
m2_000

Разделы
Разделы отсутствуют
первый

Тип записи
меню

С ответом

Тип текста ответа

Простой текст HTML

HTML-форматирование доступно только для информации

Файл ▾ Изменить ▾ Вид ▾ Вставить ▾ Формат ▾ Ин
 [Иконка] Формат ▾ Параграф ▾ Шрифт ▾ **B**

DELETE_SLOT_VALUES m*



Введите адрес документа

Кнопки:

Поиск записи в базе знаний

Значения слотов:

Ответ REDIRECT_TO_INTENT m2_1

Тип текста ответа

Простой текст HTML

HTML-форматирование доступно только для информации

Файл ▾ Изменить ▾ Вид ▾ Вставить ▾ Формат ▾ Ин
 [Иконка] Формат ▾ Параграф ▾ Шрифт ▾ **B**

REDIRECT_TO_INTENT m2_1

Приложение 4. Реализация интеллектуального бота

Для автоматического ответа на вопросы Бот может применять технологии искусственного интеллекта (глубокое обучение). Для поиска ответа Бот использует формулировки вопросов в статьях Базы знаний, разметку истории диалогов и статистику выбора статей из БЗ при работе Оператора.

Базой для обучения чат-бота служат:

1. Примеры вопросов, которые используются в статье, вкладка Примеры вопросов. Для лучшего обучения требуется введение, как можно большего числа вопросов в различных формулировках, при этом не требуется введения вопросов в различных падежах и числе. Количество вопросов не ограничено.

2. Разметка для обучения классификатор накапливается в базе знаний. Платформа предоставляет несколько способов сбора разметки и последующего управления разметкой:

2.1 Размеченные сообщения можно добавить из функционала Похожие вопросы. Система анализирует все ведущиеся диалоги на платформе и может предложить в качестве похожих вопросов подходящие с точки зрения искусственного интеллекта вопросы клиента, которые были заданы в диалогах для данной статьи. Они находятся на вкладке " Похожие вопросы" в статье. Администратор может выбрать подходящие с его точки зрения вопросы из всех предложенных. Отмеченные вопросы можно затем увидеть во вкладке "размеченные сообщения", и в дальнейшем они будут использоваться для обучения бота. Постепенное расширение (обучение) этого раздела ускоряет процесс получения ответа на вопросы Обратившегося и увеличивает процент вероятности правильности ответа.

2.2. Автоматическое накопление разметки для классификаторов в процессе работы операторов и использования ими статей базы знаний для ответов. Автоматически при работе Оператора, если Оператор использует для ответов статьи БЗ (ответ с помощью "самолетика")

2.3. Инструмент разметки истории диалогов. Ручная разметка сообщений в истории обращений (главное меню "Поиск обращений" -> функция "Автоответ").

Для запуска переобучения чат-бота необходимо в главном меню «База знаний» открыть вкладку «Управление классификаторами». В разделе формы с постфиксом «_wmd/_logreg/_knn/_ggu» (если классификаторы не видны нажать «Обновить информацию о классификаторах») справа нажать «Переобучить классификатор». Надпись «Работает» изменяется на красную надпись «Обучается», а затем, по окончании переобучения, вновь изменяется на «Работает» зеленого цвета (классификатор переобучился).

Пример обучения чат-бота с классификатором WMD

Как правильно обучать чат-бота?

Например, чтобы для каждого из таких вопросов были разные статьи и бот не путался?

1. Оформить кредитную карту.

2. Кредитные карты.
3. Подобрать кредитную карту.

Чтобы бот реагировал на ключевые слова и понимал, что от него хотят.

Также как сделать так, чтобы чат-бот не реагировал на 100% точные формулировки из Примеров вопросов? Например, есть обученная статьи с примерами вопросов типа как "Как оформить кредитную карту?", но если человек введет "Как я могу оформить кредитную карту?" - бот не находит нужного ответа, хотя 80% слов совпадает.

Ответ:

Главная причина, по которой бот путается в ответах - это дубликаты - те фразы, которые бот считает одинаковыми.

Допустим, клиент задал вопрос. Wmd ищет ближайшую фразу, наиболее похожую на вопрос. Если ближайшей к вопросу фраза присутствует в двух разных интентах, то бот получит одинаковую вероятность того, что любой из этих интентов верный, и значит вернет любой из них.

Дубликаты для вопросов про кредитные карты показаны на рис.П.4.1.

| | questions | labels | lemmas |
|------|-------------------------------|------------------------------------|--------------------------------|
| 342 | кредитная карта с кешбек | natural_cards_creditcard_owner | кредитный_A карта_S |
| 1922 | своя кредитная карта | owner_credit_card2 | кредитный_A карта_S |
| 1923 | кредитная своя карта | owner_credit_card2 | кредитный_A карта_S |
| 2112 | Вопрос по кредитной карте. | natural_cards_creditcard | кредитный_A карта_S |
| 2114 | выбрать кредитную карту | natural_cards_creditcard | кредитный_A карта_S |
| 2139 | Кредитные карты | natural_cards_creditcard | кредитный_A карта_S |
| 2123 | Получить кредитную карту. | natural_cards_creditcard | получать_V кредитный_A карта_S |
| 2657 | Где получить кредитную карту? | natural_cards_creditcard_card-draw | получать_V кредитный_A карта_S |

Рисунок – П.4.1

Здесь lemmas - это то, как фразу видит бот. Заметьте, что это та же фраза, только слова в начальной форме, и некоторые выброшены.

Можно выделить 2 ключевых этапа, на которых из фразы выпадают слова: лемматизация и фильтрация лемм, отсутствующих в эмбедингах.

1. Лемматизация. На этом этапе в предложении находятся слова и приводятся в начальную форму. Затем выбрасываются некоторые части речи, такие как предлоги и местоимения. После выбрасываются стоп-слова - слова, которые обычно не относятся к сути вопроса. Например, во 2 и 3 строчке выброшено местоимение "своя". В 4 и 5 -- стоп-слова "вопрос" и "выбрать".

2. Фильтрация лемм, отсутствующих в эмбедингах. Есть словарь эмбедингов. Если слово в нем отсутствует, то бот его не видит. В этом словаре содержится большинство русских слов. Плюс, туда также добавляются специфичные термины, индивидуальные для заказчика. Слово добавляется, если прошло первый этап, то есть для него удалось подобрать лемму, и оно не отсеялось, как незначущая часть речи, и если оно содержится в датасете - в вопросах и ответах - не менее 3 раз. Например, по этому принципу было отброшено слово "кешбек" в 1ой строчке. Оно определяется как лемма "кешбек_S". Его нет в базовых эмбедингах, и оно не добавляется в словарь, так как встречается в датасете всего 2 раза.

То есть, чтобы бот не путался, должно быть как можно меньше похожих примеров в разных интентах.

Что касается фразы "Как я могу оформить кредитную карту?". Бот видит ее как 4 слова: ['мочь_V', 'оформлять_V', 'кредитный_A', 'карта_S'].

WMD точного совпадения не находит. Поэтому ищет примеры, в которых совпадает 3 слова из 4 и сравнивает расстояние между отличающимися четвертыми. Важность слов одинакова.

Поэтому для WMD хорошие варианты ближайшей фразы не только:

"Как оформить кредитную карту?" (оформлять_V кредитный_A карта_S), но и фразы из других интентов:

"Где я могу закрыть кредитную карту?" (мочь_V закрывать_V кредитный_A карта_S)

"я могу получить кредитную карту." (мочь_V получать_V кредитный_A карта_S), которые, видимо, и кажутся классификатору предпочтительнее.

Решением здесь могло бы быть добавление леммы "мочь_V" в стоп-слова, чтобы она выбрасывалась на этапе лемматизации. К сожалению, это нельзя сейчас сделать через графический интерфейс, но мы можем сделать это вручную в коде.

GRU же даже не рассматривает интент `natural_cards_creditcard_card-draw`, которому принадлежит фраза, так как он слишком мал для него. GRU классифицирует только среди интентов, у которых есть минимум 20 примеров.

Возвращаясь к вопросу:

Есть 3 статьи.

1. Оформления кредитной карты.
2. О кредитных картах.
3. Подбор кредитной карты.

Какие примеры вопросов должны быть в каждой из них, чтобы их отличать? (по 1-2 примера достаточно).

Выход - делать примеров больше и разнообразнее.

Главное, чтобы леммы, которые остаются после препроцессинга, для разных интенгов не пересекались.

Леммы = Все слова - стоп-слова - слова_без_эмбедингов

То есть если убрать стоп-слова и слова_без_эмбедингов из предложения, то предложения из разных интенгов все еще должны отличаться.

Со словами без эмбедингов справиться легко. Если добавляете пример с необычным, редким словом, которого может не быть в словаре, то нужно создать минимум 3 примера с его использованием. Также следует иметь в виду, что у таких слов могут также возникнуть проблемы с приведением в начальную форму, то есть хорошо бы добавить по 3 примера на каждую форму слова, которую хотите использовать.

Со стоп-словами главное обращать внимание, чтобы стоп-слова не были ключевыми в примере. Если в одном интенге "кредитная карта", а в другом "вопрос о кредитной карте", то будут пересекаться, так как здесь стоп-слово "вопрос" - ключевое - единственное, которым отличаются.

Или наоборот, бывают проблемы, когда какое-то слово, которое можно было бы выбросить как стоп-слово не выбрасывается. Как например "могу". Тут, мне кажется, лучше всего, если добавляется пример с таким незначащим словом в какой-то интенге, то нужно добавить аналогичный пример с этим словом в каждый похожий интенге.

В итоге примеры для 3 статей:

- 1.1 Оформление кредитной карты.
- 1.2 Оформление кредитки.
- 1.3 Оформить кредитную карту.
- 1.4 Оформить кредитку.
- 1.5 Как я могу оформить кредитную карту.
- 1.6 Как я могу оформить кредитку.

- 2.1 О кредитных картах.
- 2.2 Кредитная карта.
- 2.3 кредитка.

- 3.1 Подбор кредитной карты.
- 3.2 Подбор кредитки.
- 3.3 Подобрать кредитную карту.
- 3.4 Подобрать кредитку.
- 3.5 Как я могу подобрать кредитную карту.
- 3.6 Как я могу подобрать кредитку.

Приложение 5. Ограничения Markdown по каналам Платформы

| | Telegram | Viber | Whatsapp - InfoBIP | ВКонтакте | Facebook Messenger | Одноклассники | Яндекс. Диалог и | Skype | WhatsApp - Streamtel | WhatsApp - MFMS |
|---|-------------------------|-----------------|--------------------|-----------------|----------------------------|---------------|------------------|---------------------------------------|----------------------|-----------------|
| Форматирование текста (со стороны бота): | <i>markdown or html</i> | <i>markdown</i> | <i>markdown</i> | <i>markdown</i> | <i>markdown (web only)</i> | | | Bot publishing closed (max 100 users) | <i>markdown</i> | <i>markdown</i> |
| жирный | + | + | - | - | + | - | - | | + | + |
| курсив | + | + | - | - | + | - | - | | + | + |
| подчеркивание | - | - | - | - | - | - | - | | - | - |
| зачеркнутый | + | + | - | - | + | - | - | | + | + |
| цвет | - | + | - | - | - | - | - | | - | - |
| размер текста | - | - | - | - | - | - | - | | - | - |
| моноширинный (код) | - | - | - | - | - | - | - | | - | - |
| ссылка (надпись со ссылкой внутри) | - | + | - | - | - | - | - | | | |
| список | - | - | - | - | - | - | - | | - | + |
| Емоji | + | + | - | + | + | + | - | | | |
| Кнопки | | | | | | | | | | |
| Кнопки | + | + | + | + | + | + | - | + | | Через шаблоны |
| Перенос текста на новую строку | - | + | - | - | - | - | - | | - | + |
| Деление кнопок на строки | + | + | - | + | - | + | - | | - | - |



| | | | | | | | | | | |
|---|---|---|----------------------|---|-----------|----------------------|---|--|----------------------|---|
| | | | | | | (в новом интерфейсе) | | | | |
| Цвет кнопок | - | + | - | + | - | + | - | | - | - |
| Кнопка с переходом на URL | + | + | - | + | +(button) | + | - | | - | + |
| Глобальные кнопки (под клавиатурой, не привязанные к сообщению) | + | + | - | + | - | - | - | | - | - |
| Емоѝ в кнопках | + | + | - | + | + | + | - | | - | - |
| Картинки в кнопках | - | + | - | - | + | - | - | | - | - |
| Лента/карусель | | | | | | | | | | |
| Карусель | - | - | - | - | - | - | - | | - | - |
| Картинка | - | - | - | - | - | - | - | | - | - |
| Кнопки | - | - | - | - | - | - | - | | - | - |
| Заголовок | - | - | - | - | - | - | - | | - | - |
| Текст | - | - | - | - | - | - | - | | - | - |
| Форматирование текста | - | - | - | - | - | - | - | | - | - |
| Цвет кнопки | - | - | - | - | - | - | - | | - | - |
| Кнопка с URL | - | - | - | - | - | - | - | | - | - |
| Типы сообщений | | | | | | | | | | |
| Номер телефона без дополнительного запроса | - | - | - | - | - | - | - | | + | + |
| Имя без дополнительного запроса | + | + | (не всегда доступно) | + | + | - | - | | (не всегда доступно) | + |
| Приветствие при первом входе в чат | + | + | - | + | + | - | - | | | |



| | | | | | | | | | | |
|---|---|---|---|---------------|---|---|---|---|---|---|
| Приветствие при каждом входе в чат | - | - | - | - | - | - | - | | | |
| Текстовое сообщение | + | + | + | + | + | + | + | + | + | + |
| файл | + | + | + | + | + | + | - | + | + | + |
| картинка (JPEG, PNG, GIF) | + | + | + | + | + | + | - | + | + | + |
| видео от пользователя | + | + | - | - | + | + | - | | + | + |
| голосовое сообщение от пользователя | + | - | - | + | + | + | - | | + | + |
| Запросить email пользователя | - | - | - | - | - | - | - | | | |
| Запросить номер телефона пользователя | - | - | - | - | - | - | - | | - | - |
| геопозиция пользователя (+возможность запросить) | - | + | - | API позволяет | - | - | - | | - | - |
| отправка точки на карте пользователю | - | - | - | API позволяет | - | - | - | | | |
| стикер пользователю | - | - | - | API позволяет | - | - | - | | | |
| ссылка на URL пользователю (специально оформленная) | - | - | - | API позволяет | - | - | - | | | |
| Подборка товаров (описываются не в CraftTalk) | - | - | - | API позволяет | - | - | - | | | |
| User is typing | - | - | - | - | - | - | - | | | |



| | | | | | | | | | | |
|----------------------|---|---|---|---|---|---|---|---|--|--|
| Read confirmation | - | - | - | - | - | - | - | | | |
| Фотография оператора | - | - | - | - | - | - | - | - | | |

**Лист регистрации изменений**

| № | Описание изменения | Раздел документа | Изменение внес | Дата |
|----------|---------------------------|-------------------------|-----------------------|-------------|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |