

Разработка агента анализа текста в вопросно-ответной системе ТВИН

Аксенов К.А.

1. Общая информация о вопросно-ответной системе «ТВИН»

Вопросно-ответная система (ВОС) «ТВИН» предназначена для разработки агентов анализа текста, чат-ботов и прикладных интеллектуальных подсказчиков. ВОС «ТВИН» разработана ООО «Уралинновация». Вопросно-ответная система «ТВИН» имеет веб-ориентированный интерфейс и доступна по адресу: <https://twin24.ai/>

Перед началом работы необходимо самостоятельно зарегистрироваться в вопросно-ответной системе «ТВИН». В данном документе приведен пример разработки прикладного агента анализа текстовой информации для деканата.

(Важно !!!) При регистрации и работе с ВОС ТВИН вы работаете с полноценной системой, которая может быть использована для ведения бизнеса (взаимодействия с клиентами и продвижения продукции и услуг), функционал описанный в лабораторной работе, направлен на изучение low/no-code интеллектуальных платформ, ориентированных на анализ текстовой информации, применение нейронных сетей и методов анализа речи и текста. Важно понимать, что функционал ВОС ТВИН роботов обзыва, синтеза и анализа речи, а также коммуникации с абонентами и другими внешними системами (с которыми ТВИН может взаимодействовать) является платным.

2. Создание агента в вопросно-ответной системе «ТВИН»

После регистрации в системе ТВИН, необходимо создать нового агента (для своей предметной области), кнопка «Добавить (Создать агента)» на вкладке «Агенты NLU» (рис. 1), далее переходите к настройке свойств агента.

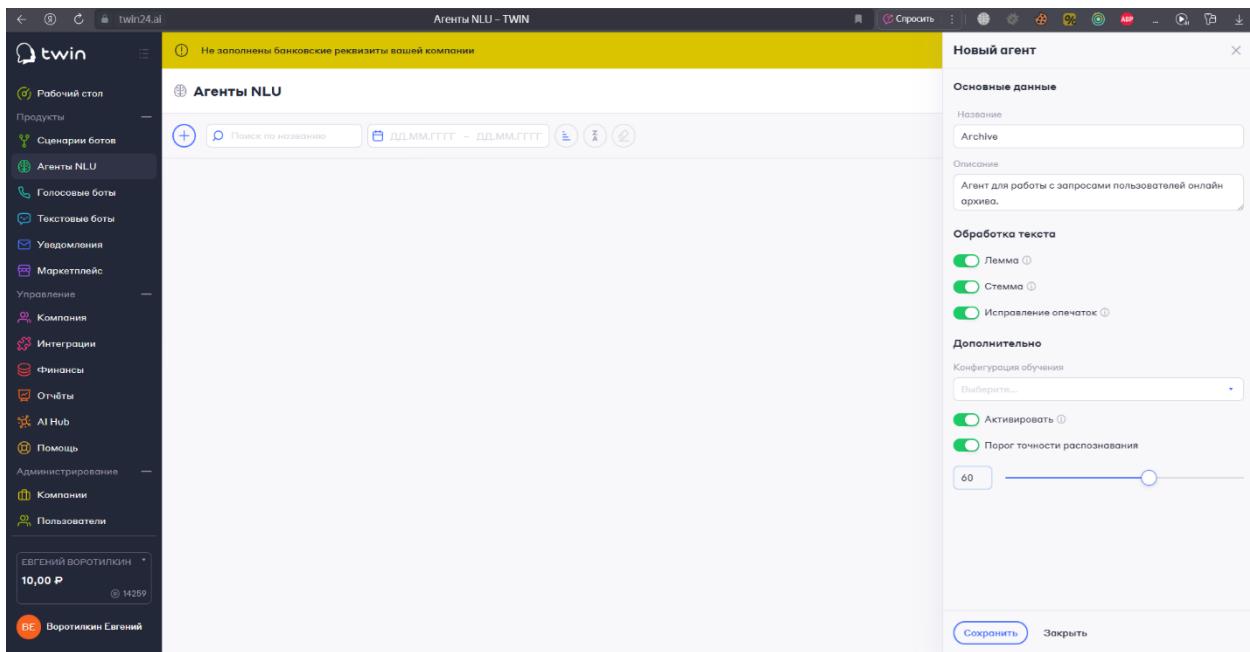


Рис. 1. Главное окно ВОС «Твин»

2.1. Настройка свойств агента анализа текста

В окне настройки свойств агента рис. 2 обязательно надо задать "Конфигурация обучения" (соответствующий язык) и один или несколько видов нормализации. ВОС ТВИН поддерживает разные языки (лабораторную можно выполнить для любого из поддерживаемых языков). В дальнейшем, можно будет корректировать настройки агента.

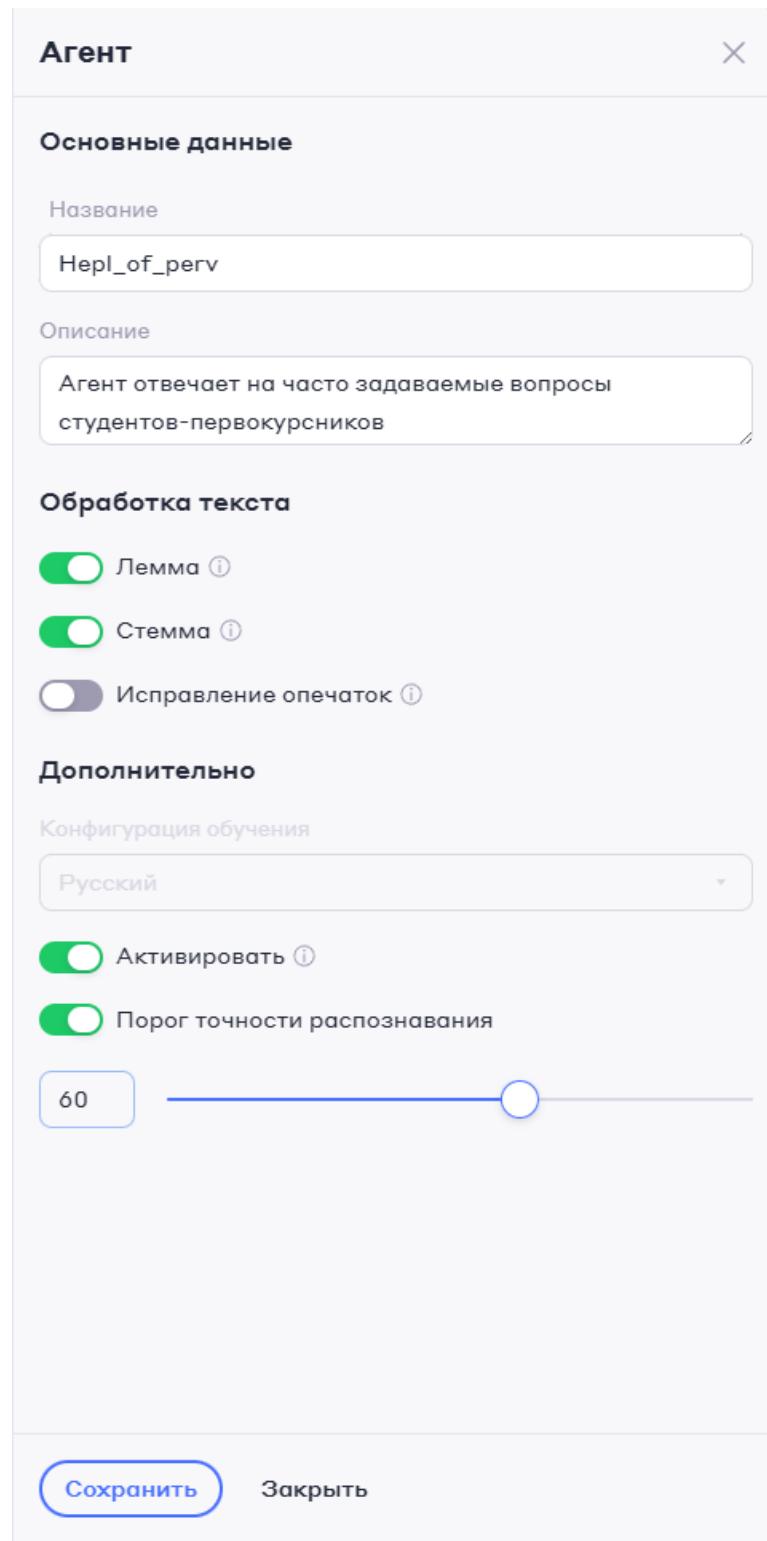


Рис. 2. Настройка свойств агента «Помощник первокурсника»

Описание: виртуальный помощник для студентов (1 курса и не только): отвечает на частые вопросы о расписании, зачетах, общежитии, библиотеке и т.д..

В документе приведены результаты двух близких по содержанию и пересечению информации агентов анализа текста «Помощник первокурсника» и «Библиотека \ Архив».

2.2. Разработка тестового диалога (сценария). Подготовительный этап.

Перед разработкой агента анализа текста рекомендуется составить текст тестового диалога (сценария). Содержащий не менее 10 вопросов и 10 ответов по вашей предметной области.

Пример диалога для предметных областей «Помощник первокурсника» и «Библиотека \ Архив» показан ниже:

Пример диалога:

- 1) Чат-бот:** Добрый день! Что Вас интересует? Чем могу помочь?
- 2) Пользователь:** Где находится библиотека?
- 3) Чат-бот:** ул. Мира д. 19 (главный учебный корпус, библиотека)
- 4) Пользователь:** Где находится деканат РТФ (Института радиоэлектроники и информационных технологий)?
- 5) Чат-бот:** Деканат РТФ находится по адресу ул. Мира д. 32 второй этаж: Р-217 (деканат бакалавриата), Р-219 (деканат магистратуры).
- 6) Пользователь:** Деканат справки выдает?
- 7) Чат-бот:** Да, деканат справки выдает. Для Вашего удобства заказать нужную справку Вы можете в личном кабинете. Как она будет готова – вам придет сообщение в личный кабинет и на электронную почту. После этого вы ее можете забрать в деканате.
- 8) Пользователь:** Как можно записаться в читальный зал?
- 9) Чат-бот:** В читальный зал можно записаться или пройдя по адресу ул. Мира д. 19 (главный учебный корпус, библиотека) или в личном кабинете «Библиотека \ Запись в читальный зал».
- 10) Пользователь:** Спасибо большое.

11) Чат-бот: Спасибо Вам за обращение! Есть ли еще вопросы?

12) Пользователь: Нет, до свидания.

13) Чат-бот: До свидания.

Разработку агента и выделение намерений и сущностей можно выполнить на основе анализа тестового диалога, а после разработки агента выполнить его тестирование на основе данного диалога.

Далее (в подразделах 2.3-2.5) описан короткий цикл (в процессе разработки агента он повторяется: 1-2) добавляем новые намерения и сущности (в каждом из них добавляем примеры фраз, содержащих примеры намерений и сущностей), 3) проводим обучение, 4) тестируем агента).

2.3. Добавление намерений

Намерение можно определить, как смысл сказанного, т.е. что имел ввиду пользователь, произнеся определенную фразу. Например, фраза «Добрый день! Хочу сделать заказ пиццы» несет 2 намерения: 1) приветствие, 2) заказ еды.

После у нас появляется наш агент (см. рис. 2), нажимаем на его название, у нас открывается вкладка намерения (рис. 3.). Далее необходимо создать намерение, нажимаем на кнопку и создаем (рис. 4.). Получается такой список (рис. 5.).

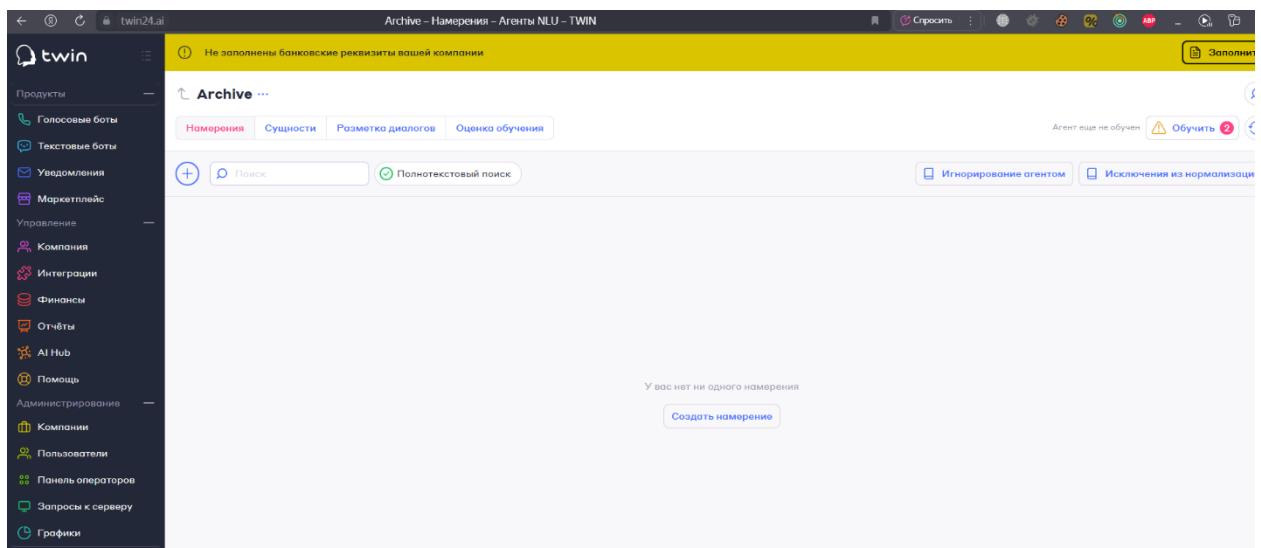


Рис. 3. Окно с разделами

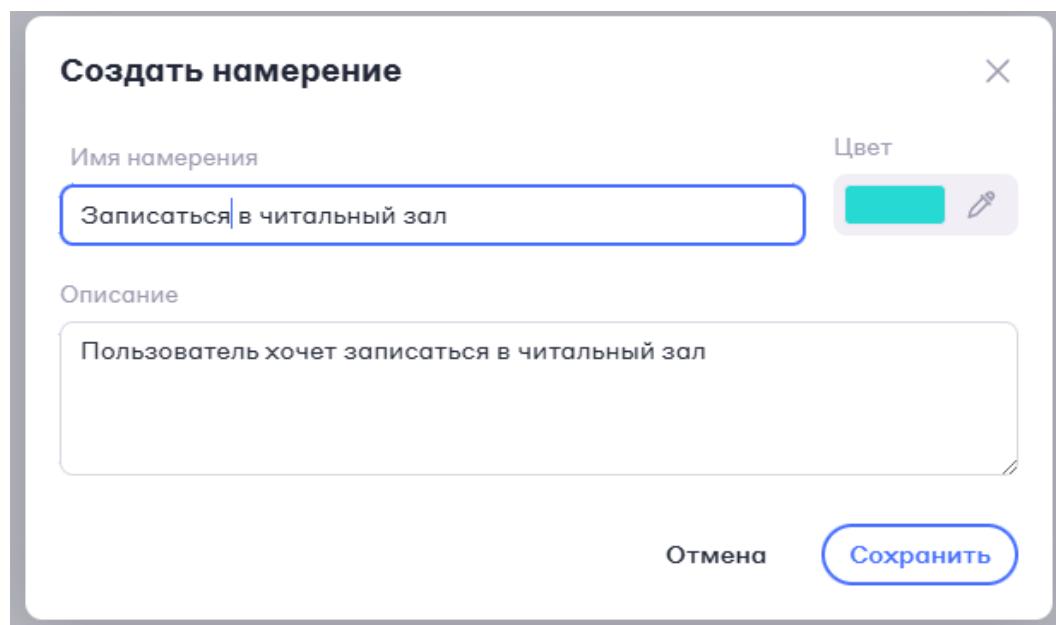


Рис. 4. Создание намерения

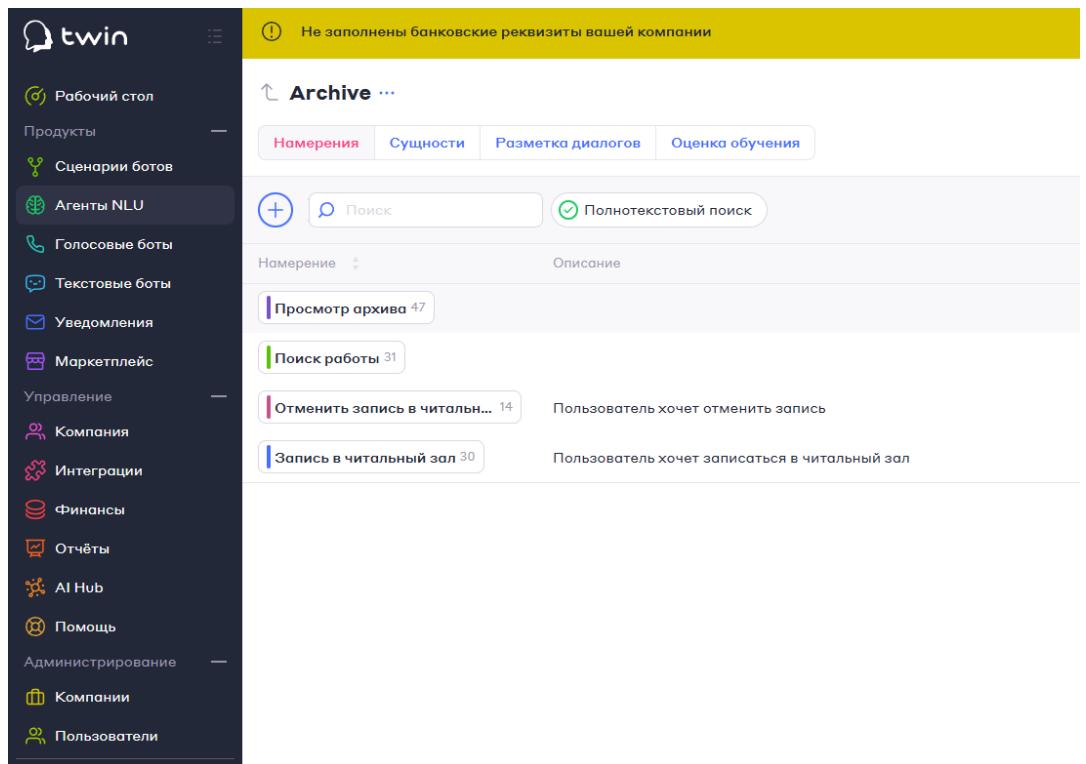


Рис. 5. Окно с частью созданных намерений

Намерения для данной предметной области представлены в Таблице 1.

Таблица 1 – Намерения и примеры фраз

№	Название намерения	Примеры фраз
1	расписание занятий	Когда у меня пары? Покажи расписание. Что завтра?
2	общежитие	Как попасть в общагу? Куда заселяться?
3	стипендия	Когда стипендия? Как её получить? Какие документы нужны?
4	деканат	Где находится деканат? Какой у меня куратор?
5	библиотека	Как взять книгу? Где библиотека? Есть ли электронный каталог?
6	записаться в читальный зал	Хочу записаться в читальный зал
7	вакансии	Какие есть вакансии?

Введенные намерения в ВОС ТВИН показаны на рис. 6.

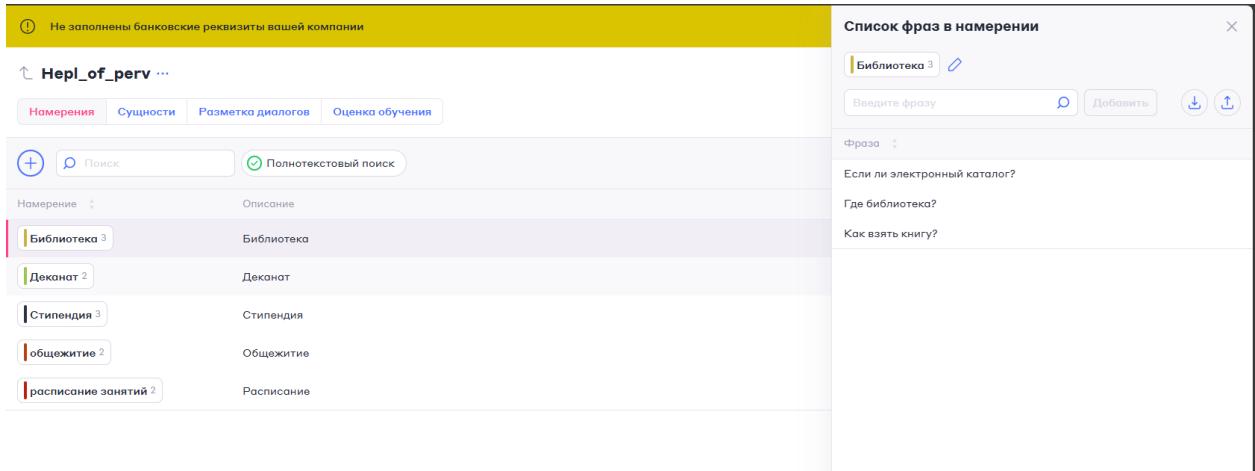


Рис. 6. Намерения агента «Помощник первокурсника» и примеры фраз

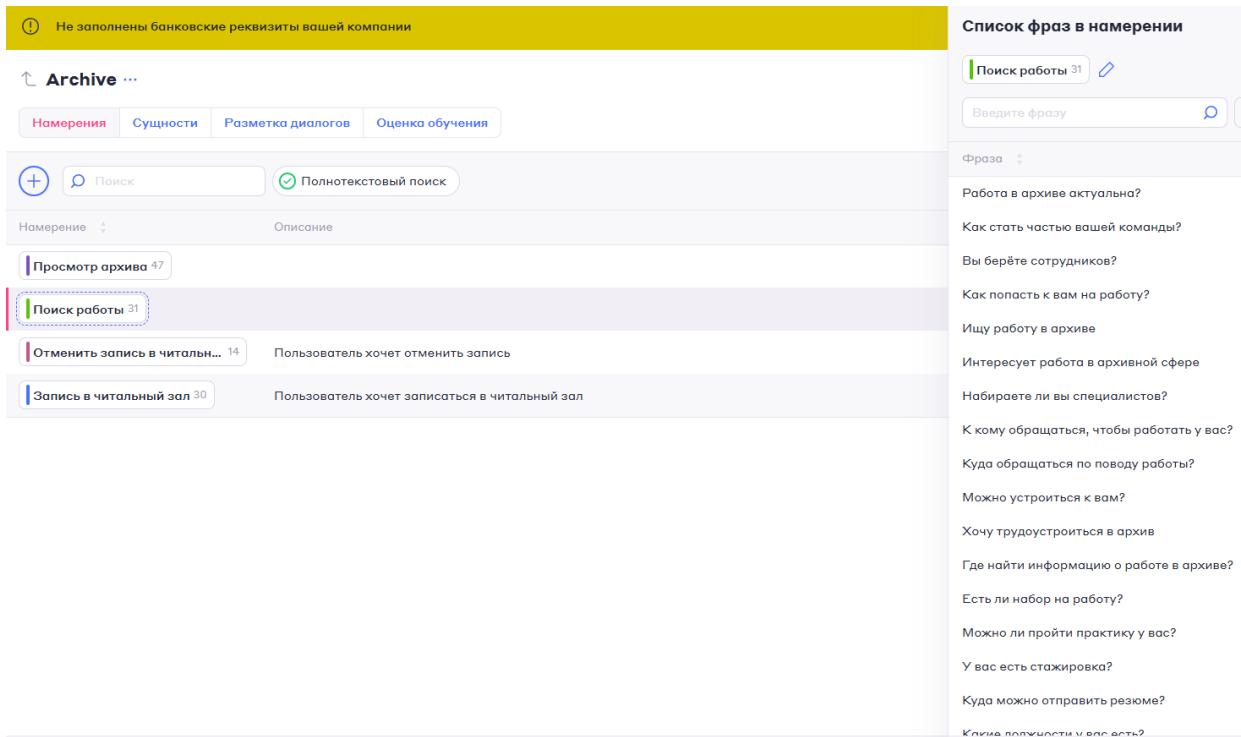


Рис. 7. Намерения агента «Библиотека \ Архив» и примеры фраз

Намерения используются в ВОС ТВИН для решения задачи классификации фразы (запроса или ответа) введенной пользователем в чат-бот, ВОС или произнесенной роботу обзвона. Эта информация используется для принятия решения о переходе на следующий шаг диалога и подготовки вопроса\ответа ботом.

2.4. Описание именованных сущностей

Следующим шагом при разработке агента анализа текста является описание сущностей предметной области.

Именованной сущностью считается слово или словосочетание, предназначенное для конкретного, вполне определённого предмета или явления, выделяющее этот предмет или явление из ряда однотипных предметов или явлений.

Для предметной области «деканат \ помочь первокурснику» часть сущностей представлены в таблице 2.

Таблица 2 – Сущности предметной области

№	Название сущности	Примеры значений
1	локация (местоположение)	деканат, общежитие, библиотека, читальный зал
2	дата	сегодня, завтра, понедельник
3	документ	студенческий, справка, заявление
4	вакансия	Архивист, методист, тьютор

Далее переходим на вкладку «Сущности» и добавляем сущности предметной области аналогично намерению. К каждой сущности добавляем примеры фразы (примеры сущностей) рис. 8.

Не заполнены банковские реквизиты вашей компании

← Непримечательны ...

Намерения Сущности Разметка диалогов Оценка обучения

+ Поиск

Имя сущности Описание

• документ 3 Документ

• дата 3 Дата

• локация 3 Локация

Просмотр сущности

Введите фразу Добавить

Фраза: Заявление, Справка, Студенческий

Описание: Документ, Дата, Локация

Рис. 8. Сущности агента «Помощник первокурсника» и примеры фраз

Пример сущностей для агента «Библиотека \ Архив» показан на рис. 9.

The screenshot shows the 'Library \ Archive' agent interface. At the top, there is a yellow banner with the message 'Не заполнены банковские реквизиты вашей компании'. Below the banner, there is a navigation bar with tabs: 'Намерения' (Intentions), 'Сущности' (Entities), 'Разметка диалогов' (Dialogue Annotation), and 'Оценка обучения' (Training Evaluation). The 'Сущности' tab is selected. On the left, there is a search bar with a plus icon for adding new entities and a magnifying glass icon for search. Below the search bar, there is a dropdown menu for 'Имя сущности' (Entity Name) and a list of entities: 'Архивный документ 17' and 'Вакансия 20'. To the right, there is a sidebar titled 'Просмотр сущности' (Entity Preview) with a search input field and a list of entity types: архивный техник, аналитик фонда, волонтер архива, практиканter, стажер, документовед, делопроизводитель, секретарь архива, ассистент архива, консультант читального зала, оператор сканирования, лаборант, научный сотрудник, руководитель отдела.

Рис. 9. Сущности агента и примеры фраз для агента «Библиотека \ Архив»

Важно! Для обеспечения релевантности поиска и решения задач классификации намерений и диагностики сущностей очень важно для каждого намерения и каждой сущности задать примеры возможных фраз. Этот этап работ с ВОС ТВИН по своей сути соответствует этапу разметки обучающей выборки при обучении нейронной сети.

2.5. Эксперименты с обучением (дообучением) агента

После добавления всех намерений и сущностей, необходимо нажать кнопку «Обучить», после чего можно будет задавать вопросы и тестировать агента (рис. 10).

Результат проверки (диагностики) фразы «Хочу записаться в читальный зал» показал следующее (рис. 10): 1) классифицировал намерение «Запись в читальный зал» 97%; 2) диагностировал сущность «organization» («читальный зал») 100%.

Результаты проверки

Намерения (1)

Запись в читальный зал 97%

Сущности (1)

100% organization

Проверка

Рис. 10. Результат диагностики фразы «Хочу записаться в читальный зал»

Результат диагностики фразы «Когда мне идти в библиотеку?» показан на рисунке 11.

Результаты проверки

Намерения (5)

Не распознано 100% | Стипендия 34% | расписание занятий 32% | Деканат 31% | Библиотека 16%

Сущности (1)

100% organization

Проверка

Рис. 11. Результат диагностики фразы «Когда мне идти в библиотеку?»

Результат диагностики фразы «Где деканат РтФ?» показан на рисунке 12.

Результаты проверки

Намерения (3)

Деканат 73%

Библиотека 40%

расписание занятий 1%

Сущности (1)

● 100% локация

≡ Проверка

Где деканат РтФ?

Введите фразу для тестирования в базе намерений и сущностей

Рис. 12. Результат диагностики фразы «Где деканат РтФ?»

Результат диагностики фразы «Библиотека работает в понедельник?» показан на рисунке 13 (в данной фразе диагностирована сущность «понедельник», диагностированные сущности можно посмотреть, наведя курсор мыши в область диагностированной сущности). Дополнительно диагностированные сущности выделяются («подсвечиваются») соответствующим цветом (цвет задается на этапе добавления сущностей в словарь агента).

Результаты проверки

Намерения (5)

Не распознано 100%

Системная: Нет

Библиотека 51%

общежитие 45%

Стипендия 6%

Деканат 6%

Сущ Найдена в слове: понедельник

● 100% time

● 100% дата

≡ Проверка

Библиотека работает в понедельник?

Введите фразу для тестирования в базе намерений и сущностей

Рис. 13. Результат диагностики фразы «Библиотека работает в понедельник?»

Результат диагностики фразы «Общежитие работает в понедельник?» показан на рисунке 14.

Результаты проверки

Намерения (2)

общежитие 72%

Стипендия 53%

Сущности (3)

● 100% time

● 100% локация

● 100% дата

 Проверка

Общежитие работает в понедельник?

Введите фразу для тестирования в базе намерений и сущностей

Рис. 14. Результат диагностики фразы «Общежитие работает в понедельник?»

Результат диагностики фразы «Когда выдадут стипендию?» показан на рисунке 15.

Результаты проверки

Намерения (2)

Стипендия 95%

расписание занятий 9%

 Проверка

Когда выдадут стипендию?

Введите фразу для тестирования в базе намерений и сущностей

Рис. 15. Результат диагностики фразы «Когда выдадут стипендию?»

Результат диагностики фразы «Во сколько завтра работает библиотека?» показан на рисунке 16.

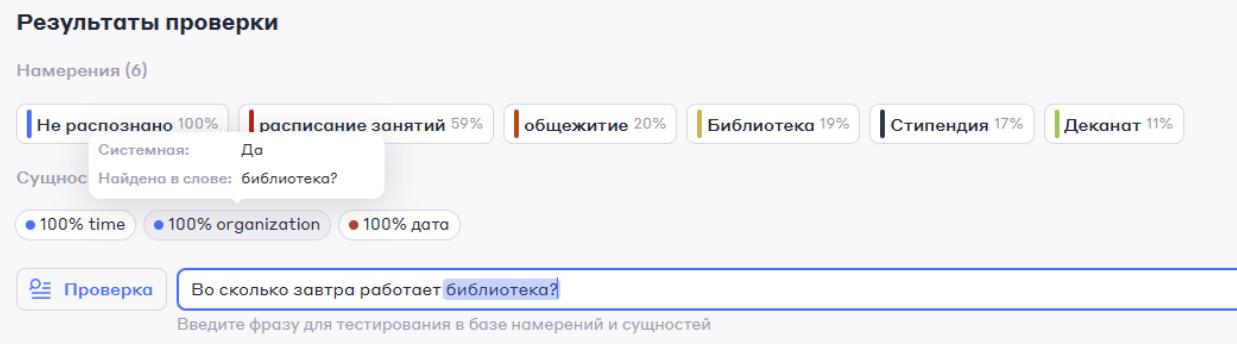


Рис. 16. Результат диагностики фразы «Во сколько завтра работает библиотека?»

Результат диагностики фразы «В деканате выдают справку?» показан на рисунке 17.

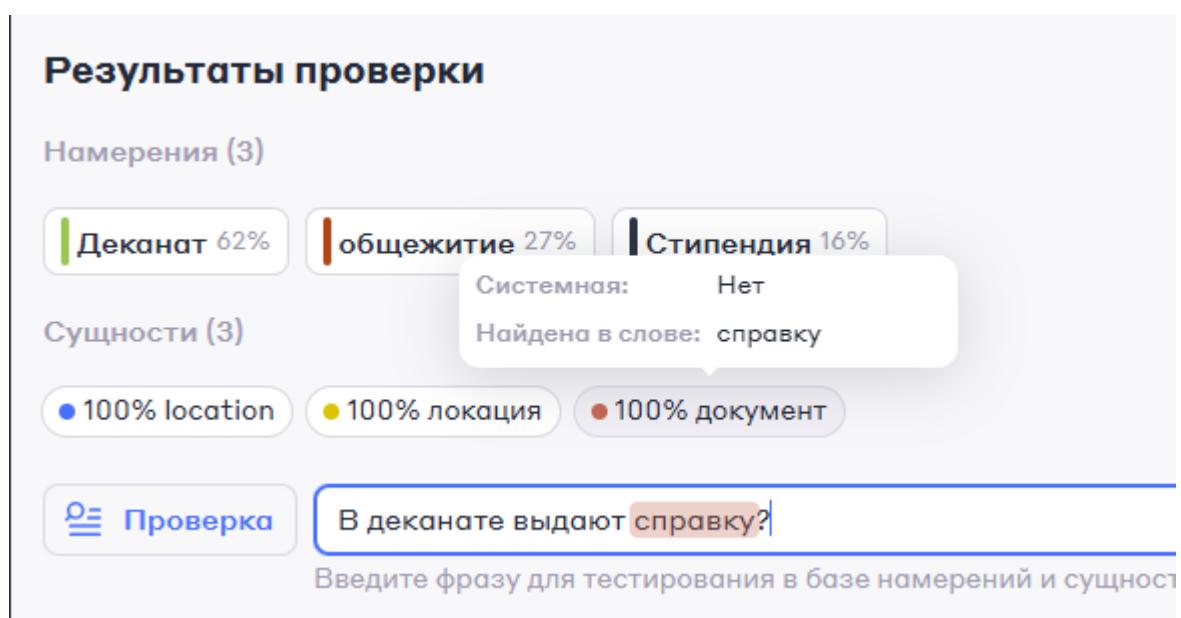


Рис. 17. Результат диагностики фразы «В деканате выдают справку?»

Результат диагностики фразы «Ищу работу архивистом» показан на рисунке 18.

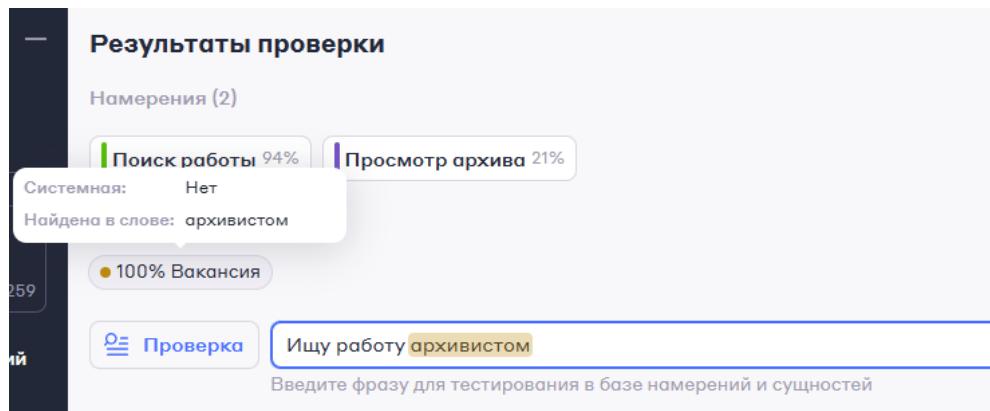


Рис. 18. Результат диагностики фразы «Ищу работу архивистом»

Результат диагностики фразы «Хочу отменить запись в библиотеку» показан на рисунке 19.

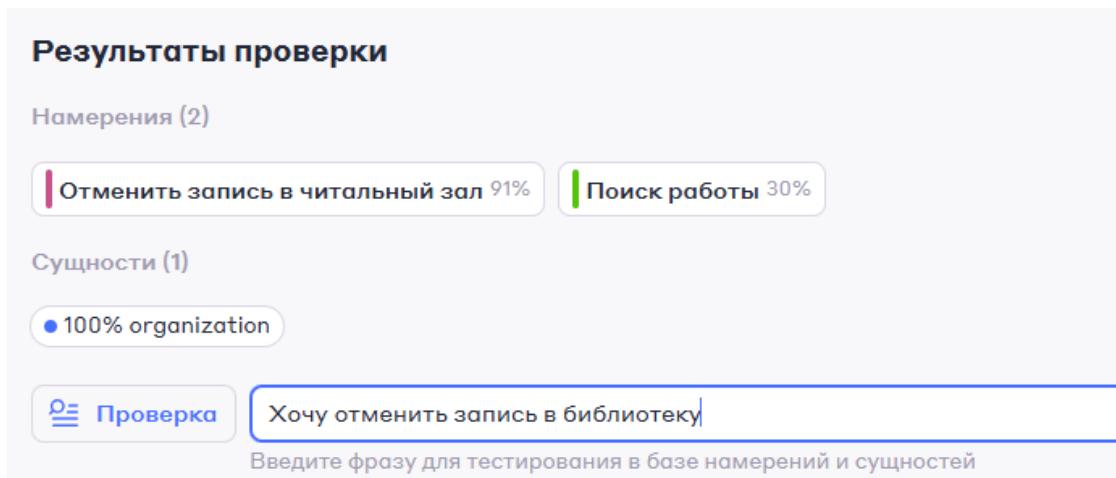


Рис. 19. Результат диагностики фразы «Хочу отменить запись в библиотеку»

Результат диагностики фразы «Как записаться в читальный зал» показан на рисунке 20.

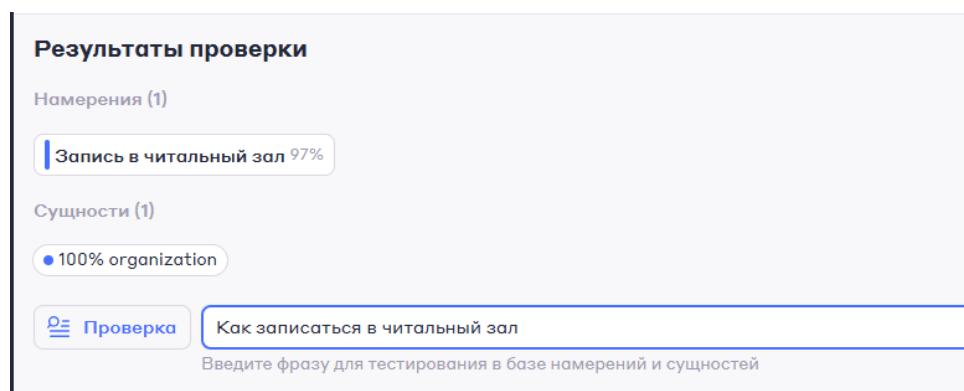


Рис. 21. Результат диагностики фразы «Как записаться в читальный зал»

Результат диагностики фразы «Где находится деканат» показан на рисунке 22.

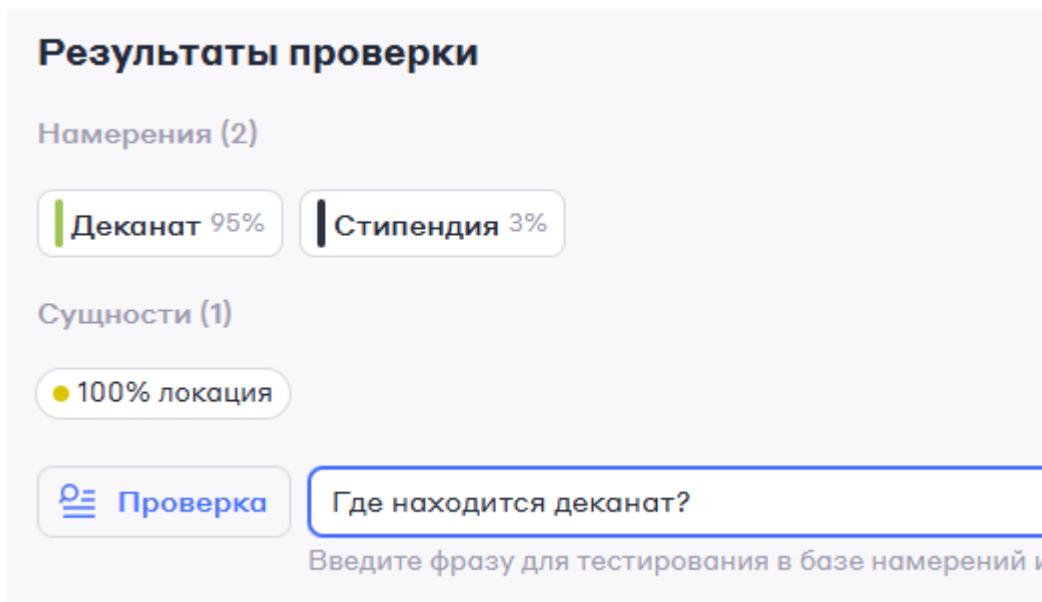


Рис. 22. Результат диагностики фразы «Где находится деканат»

Далее в свободной форме описаны примеры тестирования агента и возникшие вопросы. **После реализации регулярных выражений и ввода большего набора фраз и их разметки, качество поиска агента повысится.**

На рисунке 23 показан пример, в одной из предыдущих версий ВОС ТВИН (веб интерфейс изменился), когда не диагностируется населенный пункт «Полевской».

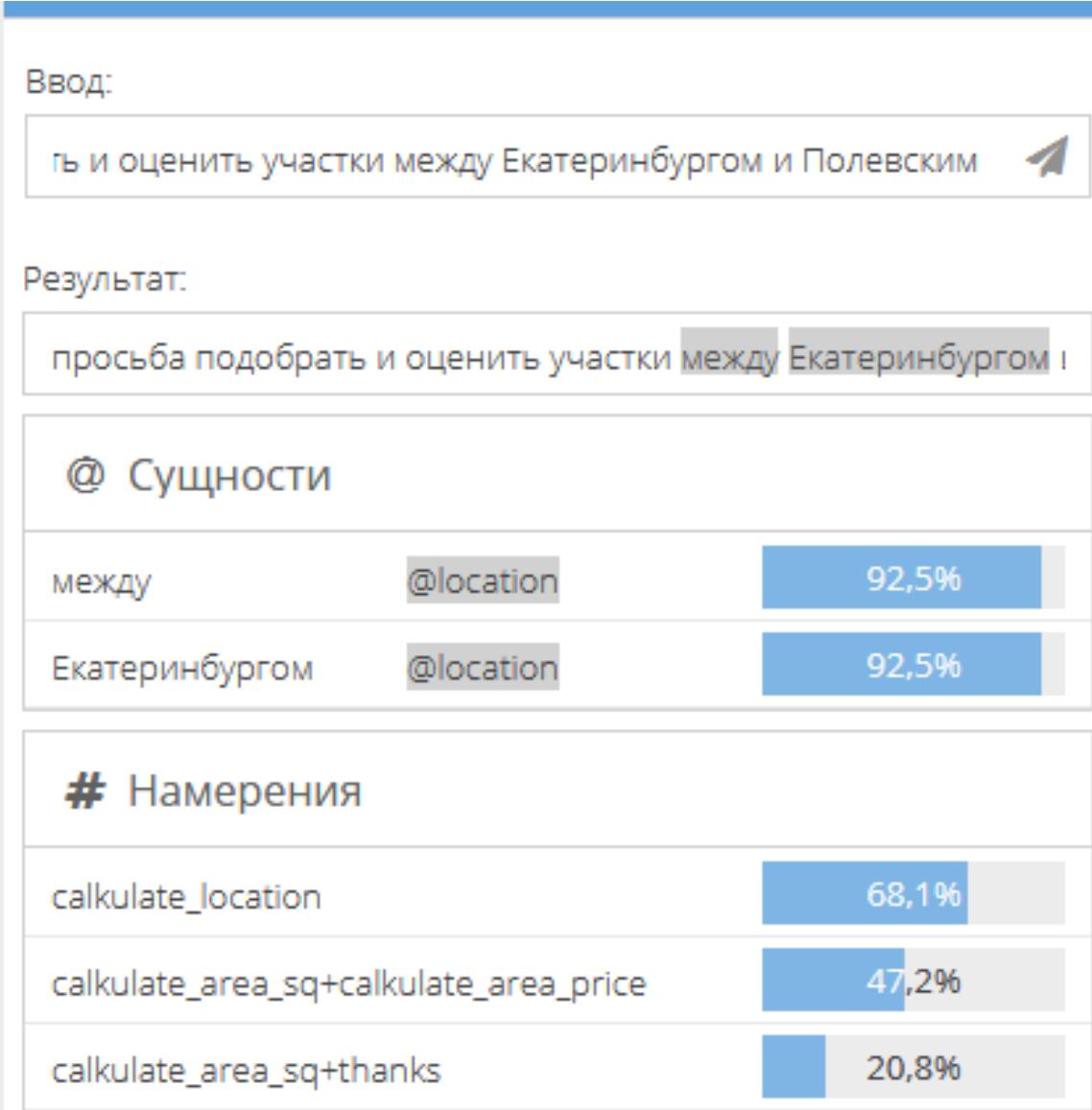


Рис. 23. Не диагностируется населенный пункт «Полевской»

На рис. 24 показана ситуация, после дополнения в обучающую выборку (добавление примеров фраз в соответствующие классы сущностей и намерений) пункта «Полевской» и дообучения агента.

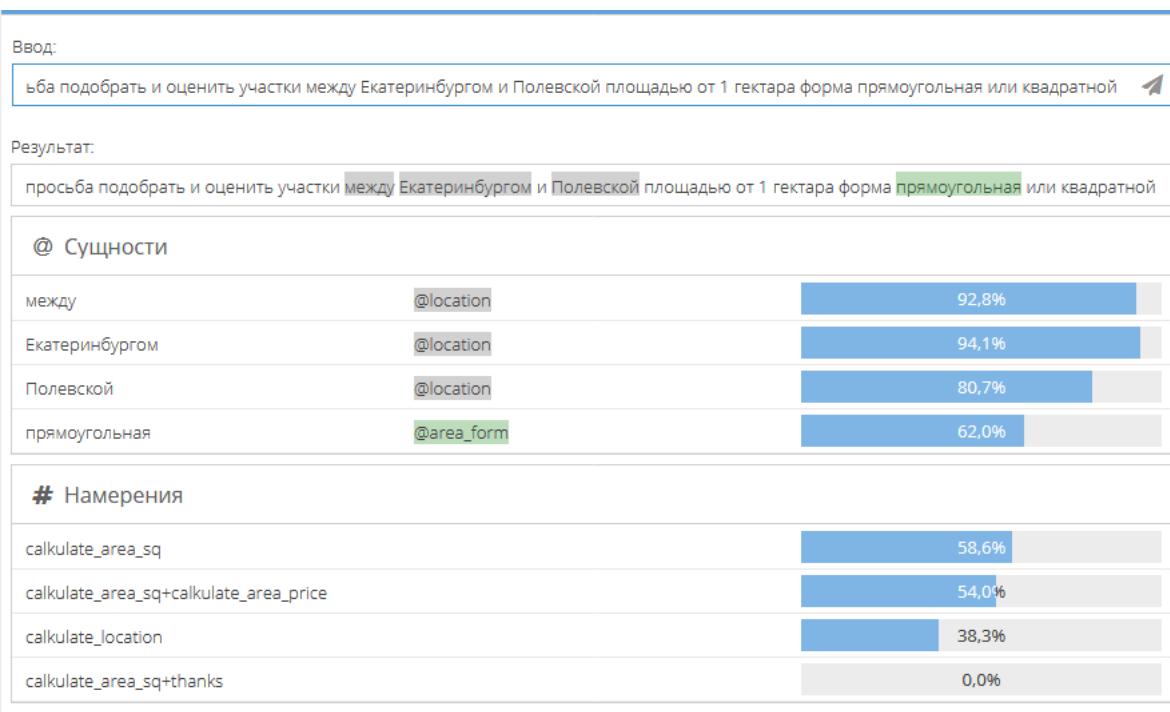


Рис. 24. Населенный пункт «Полевской» диагностируется после дообучения агента

Задание для самостоятельной лабораторной работы в вопросно-ответной системе ТВИН:

Для своей предметной области разработайте агента анализа текста в вопросно-ответной системе ТВИН. Перед началом работы в ВОС ТВИН выделите (составьте списки) для своей предметной области: 1) сценарий тестового диалога (10 вопросов и 10 ответов); 2) список намерений; 2) список сущностей. Сценарий тестового диалога должен содержать не менее 20 фраз. По окончании разработки агента анализа текста составьте отчет.

Важно!!! При оформлении отчета и вставке в отчет результатов диагностики сущностей и намерений каждой из 20 фраз (на снимаемых скринах (картинах) ВОС ТВИН) обязательно надо отображать, что искомая сущность у Вас диагностировалась (в окне диагностики нужно на нее навести мышью (появится всплывающее окно (см. рис. 13, 16-18).

Рекомендации по проверке правильности выполнения лабораторной работы в коммерческой версии ВОС ТВИН:

- 1) Отчет по лабораторной работе должен содержать результаты этапов структуризации, концептуализации и формализации знаний тестового агента для выбранной предметной области: а) списки намерений и сущностей (допустимы скрины (картинки) данных списков из ВОС ТВИН); б) сценарий тестового диалога (содержащий 10 вопросов и 10 ответов), допустимо 2 диалога с суммарным количеством фраз не менее 20.
- 2) Результаты обучения и тестирования разработанного агента на тестовом сценарии должны демонстрировать корректную диагностику намерений и в большинстве случаев корректную диагностику сущностей (допустимы некоторые ошибки диагностики сущностей (связанные с особенностью применением аппарата нейронных сетей)). Некорректная диагностика сущностей должна быть зафиксирована в отчете. В большинстве случаев некорректная диагностика сущностей может быть устранена процессом дополнительного ввода (расширения) обучающей выборки (дополнительный ввод фраз, содержащих примеры с данной сущностью, и их разметка).
- 3) Скрины тестирования агента (диагностики сущностей и намерений) должны быть хорошо читаемы.

Вопросы для самоконтроля

1. Что такое сущность?
2. Чем сущность отличается от намерения?
3. Приведите примеры сущностей.
4. Приведите примеры намерений.
5. Приведите примеры намерений, содержащие сущности.
6. Приведите примеры намерений, не содержащие сущности.

7. Что такое намерение?

8. Верно ли утверждение, что в вопросно-ответной системе ТВИН при разработке агента приходится настраивать гиперпараметры нейронной сети?

9. Верно ли утверждение, что в вопросно-ответной системе ТВИН этапу разметки экземпляров намерений соответствует этап ввода обучающей выборки?

10. Будет ли корректно работать агент анализа текста в вопросно-ответной системе Твин если будут заведены только словари сущностей и намерений?

11. Какие методы и аппараты, на ваш взгляд, используются в вопросно-ответной системе ТВИН?

12. Верно ли утверждение, что ВОС ТВИН представляет собой low-code систему для разработки роботов обзыва, чат-ботов, агентов анализа текста?

13. Верно ли утверждение, что намерения и составные намерения в ВОС ТВИН используются для решения задачи классификации ответов\запросов пользователя\абонента?

14. Верно ли утверждение, что на качество диагностики агента ВОС ТВИН в большей мере оказывает этап ввода примеров составных намерений и разметки текста?

Список литературы:

WIN — это омниканальная платформа для коммуникаций бизнеса с клиентами и для создания интеллектуальных ботов на её основе

уководство пользователя (администратора компании) Омниканальная П

arasiev A. A., Filippova M. E., Aksyonov K. A., Talantsev E. N. 2024, Application of a Scenario Based Dialog Expert System to Automation of Different Subject Areas. *AIP Conference Proceedings: book*. Vol. 3094, 190005, American Institute of Physics Inc., International Conference of Numerical Analysis and Applied Mathematics 2022, ICNAAM 2022, Heraklion, Греция, 19.09.2022. <https://doi.org/10.1063/5.0210663>

arasiev A., Filippova M., Aksyonov K., Aksyonova O., Antonova A. Using of a

Open-Source Technologies for the Design and Development of a Speech Processing System Based on Stemming Methods. IFIP Advances in Information and Communication Technology Volume 582 IFIP, 16th IFIP WG 2.13 International Conference on Open Source Systems, OSS 2020; I

ксенов К.А., Сунь Лина, Калинин И.А. Разработка системы распознавания намерений и сущностей для систем вопросов и ответов с использованием “no code” платформы “TWIN”. Инженерный вестник Д

ксенов К.А., Сунь Лина. Эволюция и современное состояние систем ответов на вопросы: технологии распознавания намерений и именованных сущностей с использованием модели BERT. Инженерный

в

е

с

т

и

к

)

Д

о

н

а

т

(

1

1

5

)

)

:)

н

С

т

р

h

t

t

p

w

w

w

i

y