

# Платформа ForQ: no-code решение для управления знаниями

Ксения Студеникина, ведущий лингвист-разработчик  
Кир Шиманович, руководитель отдела NLP  
Unistar Digital

## Чат-боты



короткие  
сообщения

Но...

автоматизация техподдержки

обработка часто  
задаваемых вопросов

длинный текст

+

сложность модерации

автоматизация техподдержки



**Базы знаний**



обработка часто  
задаваемых вопросов

длинный текст

+

сложность модерации

# Что такое база знаний?

Набор вопросов и ответов для предметной области

- Статья с вопросами-синонимами и ответом для каждой тематики
- Интерфейс для наполнения и публикации статей, отслеживания сроков актуальности и истории изменения

# Что такое ForQ?

Платформа для управления базами знаний

- Осуществляет сбор, хранение, поиск и выдачу информации из баз знаний
- Настройка каналов (голос, текст) отдельно для каждой базы знаний
- Проверка на дублирование вопросов в статьях при публикации

# Неограниченное количество баз знаний



- Проекты
- Базы знаний
- Анализ диалогов
- Отчеты
- Логи ^
- Настройки ^

10 1 из 1 < >















Создать новую

Опубликовать

Название базы	Количество статей	
Small talk	2	
Вопросы про льготы	3	
Вопросы про направления подготовки	3	
Вопросы про общежитие	10	
Вопросы в ПК МТУСИ	33	

# Неограниченное количество статей в базе знаний



Название	Эталонный вопрос	Количество фраз	Включено	Опубликовано	
Вступительные испытания	Да	66	да	да	 
Проходные баллы	Да	62	да	да	 
Платное обучение	Да	35	да	да	 
Заочное обучение	Да	27	да	да	 
Одинаковые баллы	Да	12	да	да	 
Индивидуальные достижения	Да	35	да	да	 
Сроки	Да	33	да	да	 

## От 10 вопросов-синонимов

Срок актуальности

Бесконечно

Включено

Эталонный вопрос

Как оформить целевое обучение

Вопросы синонимы

id Вопрос

1 Какие партнеры предоставляют целевое обучение?

2 Как можно подать заявку на целевое обучение?

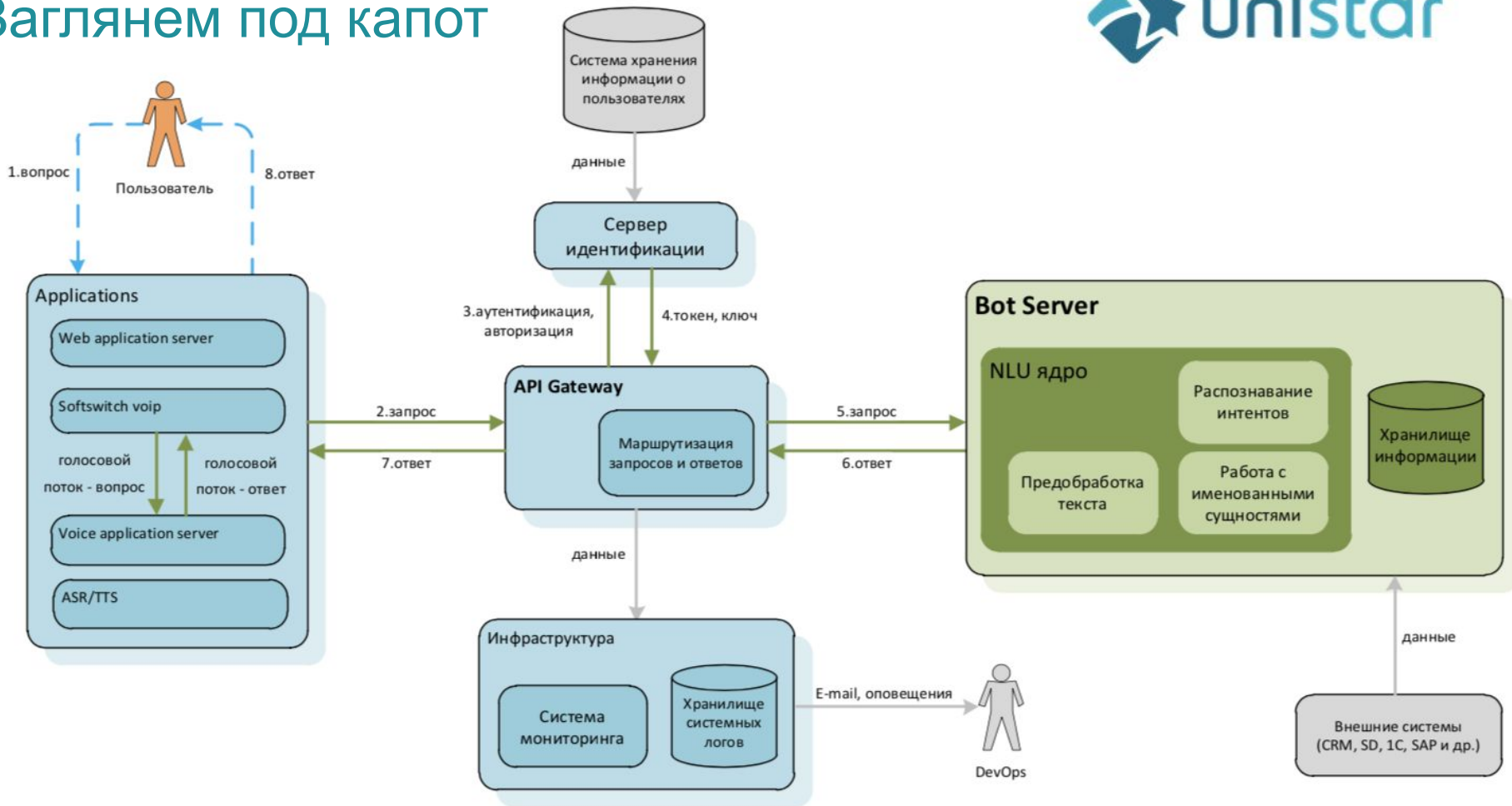
3 есть договор на целевое обучение ребенка

4 есть ли у вас возможность поступить по целевому набору

5 целевое обучение



# Заглянем под капот



## Предобработка текста

- Предобработка текста — очистка текста (удаление e-mail, номеров телефонов и т.д.), удаление стоп-слов, лемматизация.

<b>МАМА</b>	<b>МЫЛА</b>	<b>РАМУ</b>
		
<b>МАМА</b>	<b>МЫТЬ</b>	<b>РАМА</b>

## Выделение эмбеддингов

Выделение эмбеддингов — преобразование слов в вектора + добавление различных фичей.

- ✓ word2vec
  - Частеречные теги
- ✓ fasttext
  - Нет ограничения словарем

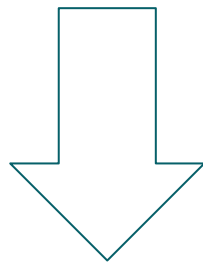
Сохраняется SLA  
меньше 1 секунды  
на ответ

- ✓ BERT
- ✓ ELMo

При возможности установить  
сервер с GPU

# Классификация

✗ Косинусное сходство — быстрое решение, но плохие метрики. Решили попробовать чуть посложнее



✓ Метод опорных векторов — пришли после тестов с другими стандартными алгоритмами ML

# Скрипты для дополнительной обработки

- Исправление опечаток
- Дедубликация статей

[← Вернуться](#) [Сохранить](#)

Наименование	<input type="text" value="Бот для приемной комиссии"/>
Требуемая точность	<input type="range" value="80"/>
Язык	<input type="text" value="ru"/>
Приоритетный алгоритм	<input type="text" value="fasttext"/>
Схожесть вопросов	<input type="checkbox"/>
Схожесть ответов	<input checked="" type="checkbox"/>
Проверка орфографии	<input checked="" type="checkbox"/>

Проект: Бот для приемной комиссии

## No-code платформа

- Понятный интерфейс по наполнению баз знаний
- Модерация базы знаний → специалист предметной области
- Первичное наполнение базы знаний → специалист по NLP

# Проект для АО «ЦентрИнформ» (ЕГАИС Росалкогольрегулирование)

ForQ как помощник для службы техподдержки ЕГАИС

- Вопросы поступают по почте → большая длина запросов
- Первичное наполнения базы знаний → кластеризация выборки обращений



# Проект для приемной комиссии вуза

ForQ для ответов на часто задаваемые вопросы











- Много вопросов в период подачи документов → разгрузить контакт-центр
- Список вопросов от приемной комиссии & примеры вопросов с Jivo-site
- В качестве диалогового движка используется JAICP



# Анализ диалогов

- Добавление нераспознанных вопросов в статьи

- Проекты
- Базы знаний
- Анализ диалогов**
- Отчеты
- Логи
- Настройки

Дата	Ответ найден	Отформатировано	Статья	Вопрос	Ответ	Процент	Проект	
26-7-2021 12:10:48	Да	Нет	Проходные баллы	с каких баллов можно участвовать?	Я виртуальный помощник приемной комиссии. Вот что я нашел по Вашему вопросу. Установлены следующие... Полный ответ	86.06%	Бот для приемной комиссии	 
26-7-2021 12:10:0	Да	Нет	Кому полагается	Кому выдается общежитие?	Я виртуальный помощник приемной комиссии. Вот что я нашел по Вашему вопросу. Общежитие предоставляет... Полный ответ	93.27%	Бот для приемной комиссии	 
26-7-2021 12:9:49	Да	Нет	Приветствие	Здравствуйте!	Здравствуйте!	100%	Бот для приемной комиссии	 
26-7-2021 12:8:15	Да	Нет	Сколько мест	где посмотреть количество мест поступающих по особой квоте? по каждому направлению	Я виртуальный помощник приемной комиссии. Вот что я нашел по Вашему вопросу. С 1 сентября 2021 года ... Полный ответ	83.28%	Бот для приемной комиссии	 
26-7-2021 12:6:51	Нет	Нет	Сроки	Добрый день! В вашем ВУЗе действительно 25 июля закончилась подача заявлений о приеме документов по ... Полный вопрос	В скором времени Вам ответит сотрудник университета.	70.19%	Бот для приемной комиссии	 

## Повторное внедрение

База знаний, обученная для одной предметной области, может быть многократно использована повторно

- Внедрение обученной базы знаний ForQ в МТУСИ (1 месяц)
- Использование уже обученной базы знаний в МАИ после небольших изменений (5 дней)



# RoadMap

- Решение в Solution Store (до конца 2021 г.)
- Добавление стоп-слов в настройках проекта (до конца 2021 г.)
- Разработка системы контроля версий и добавление возможности вернуть любую предыдущую версию (I квартал 2022 г.)
- Создание удобного интерфейса для подключения сторонних NLU-движков (I квартал 2022 г.)
- Построение предобученных баз для разных отраслей (II квартал 2022 г.)

Спасибо за внимание!