

Question Answering System

СПбАУ, МИТ, 2 курс, весна 2016

Авторы: Надежда Бугакова, Елизавета Третьякова

Руководители: Зорин Кирилл Михайлович, Калакуцкий Аркадий
Владимирович

Цель:

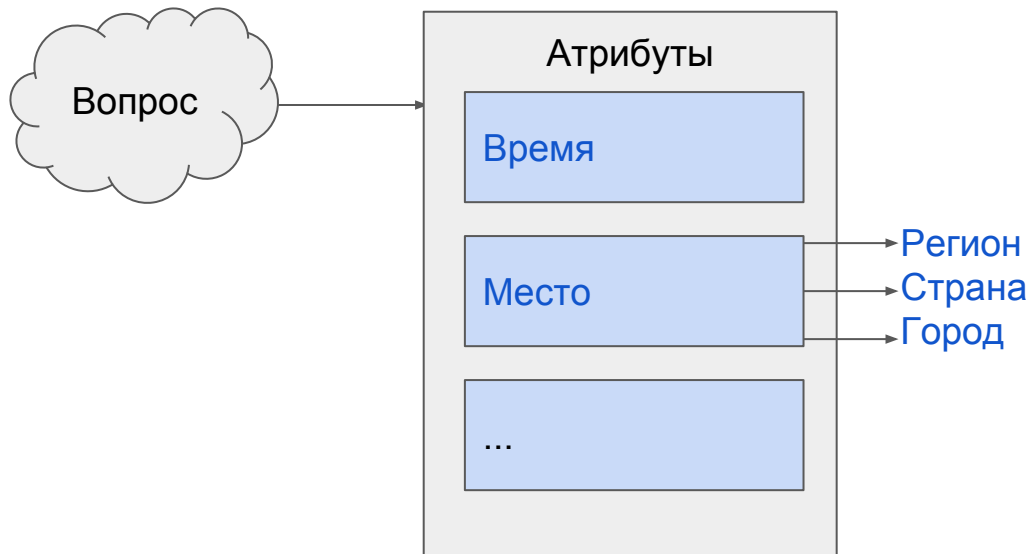
Система, получая на вход вопрос на естественном языке, пытается на него ответить, используя имеющиеся данные

Зачем это нужно:

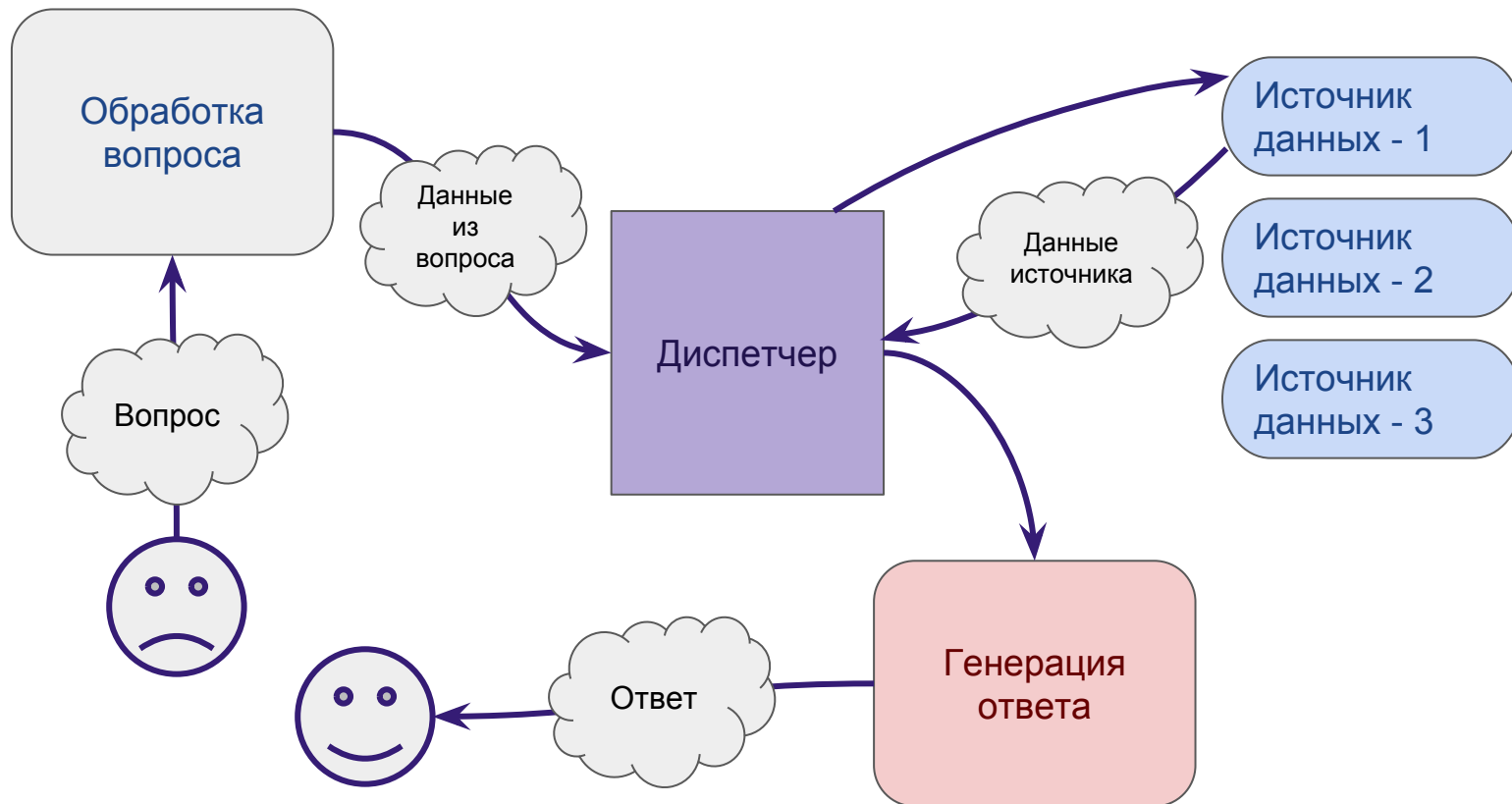
Для упрощения взаимодействия с различными источниками данных посредством создания простого и понятного единого интерфейса.

Задачи:

1. Анализ способов обработки естественного языка
2. Проектирование основных структур и понятий, используемых в системе
3. Реализация

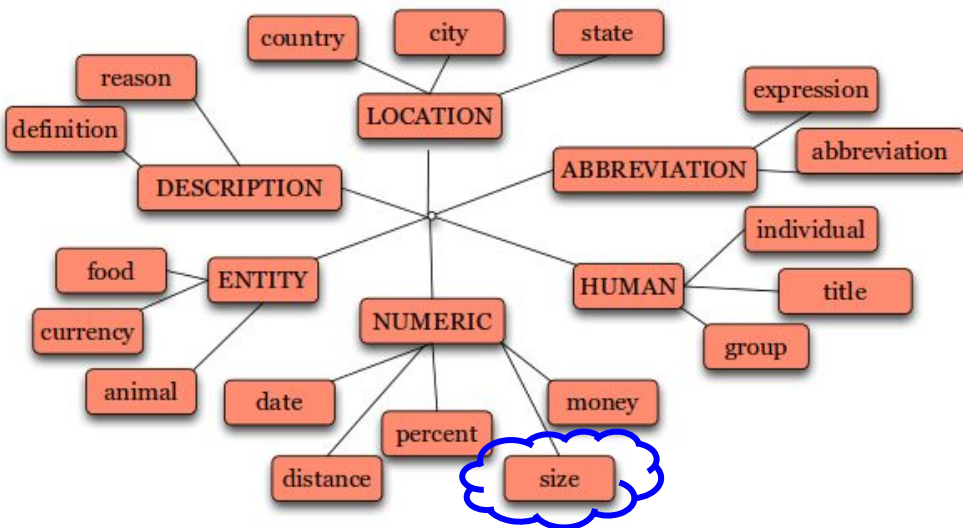


Архитектура проекта:

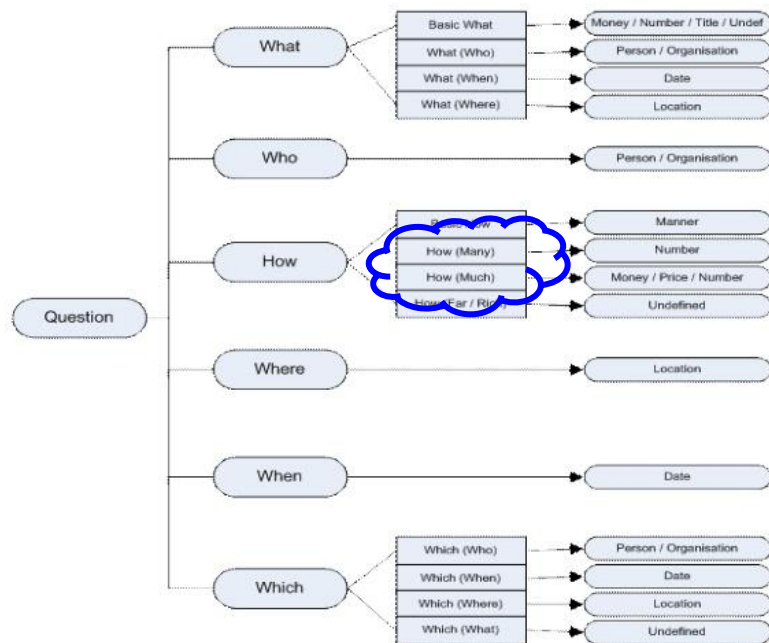


Классификация вопросов

Типы ответов:



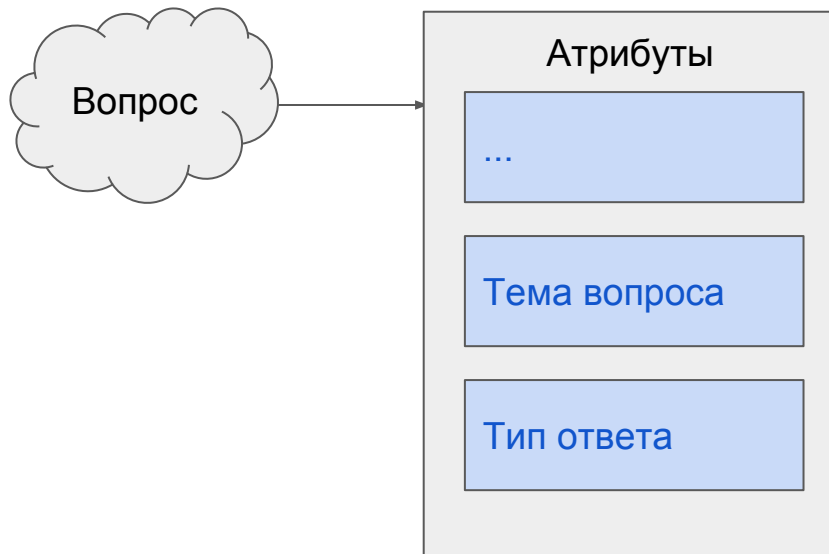
Типы вопросов:



Темы вопросов:

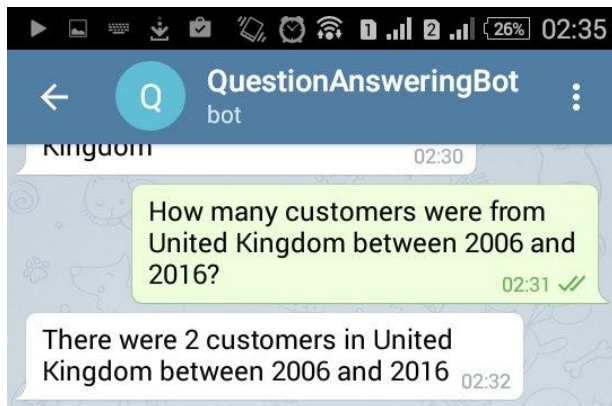
Темы вопросов, покрываемых нашей системой:

1. О количестве загрузок продукта (в т.ч. с указанием места или времени)
2. О количестве клиентов (в т.ч. с указанием места или времени)



Демонстрация (бот):

Вопрос



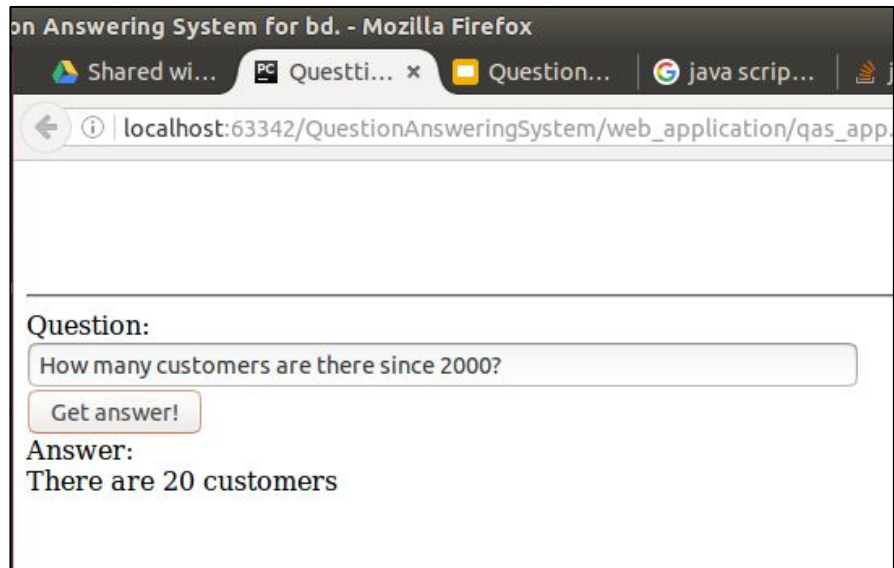
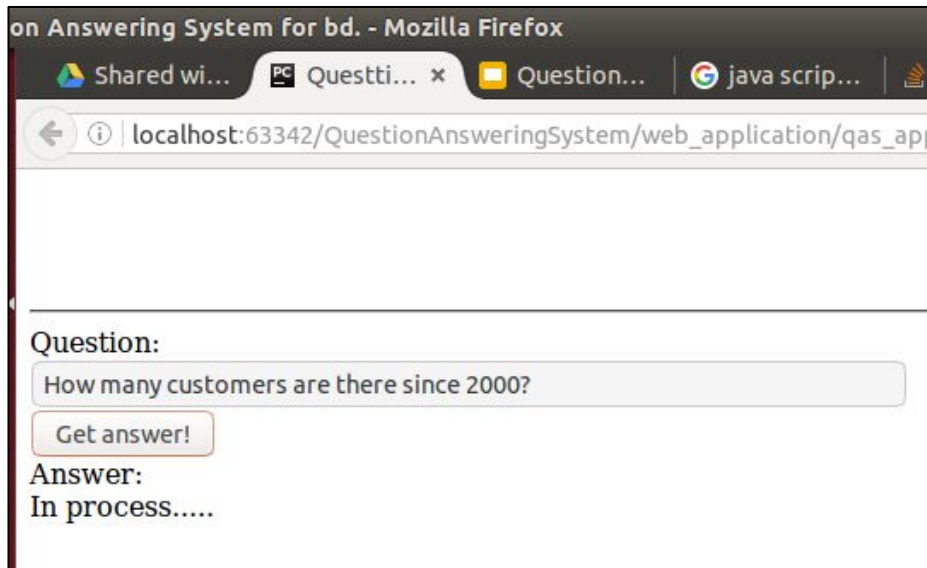
How many times the product was downloaded from Russia or Germany?

Ответ



Message

Демонстрация(локальный Web-интерфейс)



Реализовано:

1. Извлечение местоположения, времени, типа вопроса и ответа
2. Модули доступа к базам данных
3. Генерация ответа на естественном языке
4. Функциональные тесты
5. Web-интерфейс, Telegram Bot

Возникшие сложности:

- Предложения с грамматическими ошибками или опечатками
- Ограничения GeoPy.Nominatim
- Ограничения Spacy:
 - Не обнаруживает даты определённого формата
 - Не выделяет некоторые страны
 - Ресурсоёмкий

Дальнейшее развитие

1. Обработка синонимов, более широкое использование онтологий
2. Поддержка дополнительных источников данных
3. Расширение списка поддерживаемых вопросов
4. Обработка предложений с грамматическими ошибками и опечатками

Новое:

1. NLTK, Spacy
2. GeoPy
3. Работа с базами данных (использовался PostgreSQL)
4. Telegram Bot API
5. JavaScript
6. Python

Ссылка на репозиторий:

<https://github.com/LizaTretyakova/QuestionAnsweringSystem>

Спасибо за внимание!