

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

На правах рукописи

УДК 621.395.721.5:004.058.4

Анисько
Дмитрий, Геннадьевич

Системный подход к формированию FAQ-программных средств
регистрации отношения диалоговой реплики к тематической категории

АВТОРЕФЕРАТ

на соискание степени магистра техники и технологий

по специальности 1-59 81 01 «Управление безопасностью
производственных процессов»

Магистрант Д. Г.
Анисько

Научный руководитель
Шелег Валерий
Константинович,
доктор. техн. наук,
профессор

Минск 2019

ВВЕДЕНИЕ

На сегодняшний день IT-технологии развиваются с огромной скоростью. Это обусловлено наличием определенных потребностей на рынке. Потребителями IT-продуктов являются малые, средние и крупные бизнесы, а также физические лица. Одна из наиболее популярных потребностей связана с оптимизацией и ускорением обработки бизнес-процессов. Компании стараются настроить процессы таким образом, чтобы они занимали, как можно меньше времени и требовали, как можно меньше ресурсов. Один из примеров такой оптимизации – замена людей-операторов чатботами в колл центрах. Такие чатботы должны уметь отвечать на простые общие вопросы, не требующие детального анализа ситуации, а также вовлекать в диалог человека-оператора, если это необходимо.

С 2008 до настоящего времени пользуются большой популярностью кнопочные чатботы. Чтобы общаться с такими ботами пользователю необходимо нажимать на кнопки, а общение проходит по заранее определенному алгоритму.

Но кнопочные чатботы не являются интуитивно понятными для большинства людей. Поэтому в последнее время широкое использование получила технология NLP (Natural Language Processing).

NLP – это подраздел информатики и искусственного интеллекта, посвященный тому, как компьютеры анализируют естественные языки. NLP позволяет применять алгоритмы машинного обучения для текста и речи.

С использованием технологии NLP создано очень много сервисов и приложений, к которым люди привыкли и используют в повседневной жизни. К таким относятся: Apple Siri, Яндекс.Алиса, Microsoft Cortana и др. Также были создано много инструментов для упрощения разработки NLP-приложений, например, Microsoft Azure QnA Maker, Google Dialogflow и др. С помощью технологий NLP решаются и другие задачи, например, определение спама в электронной почте. Поисковые системы также используют внутри себя NLP-движки.

NLP-чатботы способны проанализировать реплику на естественном языке, определить, что человек имел в виду, и выполнить необходимое действие. Такие чатботы также способны обучаться, пополнять свои базы знаний и, в будущем, лучше выполнять свою задачу.

Целью магистерской диссертации является разработка NLP-платформы для регистрации отношения диалоговой реплики к тематической категории.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы диссертации. Для того, чтобы разработать качественный программный продукт, нужно знать потребности рынка и современные методы разработки программного обеспечения. В работе приводится анализ рынка по программному обеспечению в сфере обработки естественных языков и искусственного интеллекта. На сегодняшний день одной из основных потребностей является автоматизация бизнес процессов, минимизация человеческих трудовых ресурсов. В диссертации рассмотрены программные продукты, аналогичные разработанному, выделены достоинства и недостатки. Исходя из анализа были сформулированы требования для программного средства.

Объектом исследования является научная сфера распознавания естественных языков.

Предметом исследования являются вопрос-ответные системы распознавания естественных языков.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В начале работы проводится анализ предметной области, а именно сферы обработки естественного языка. Приводятся основные задачи и проблемы этого направления. Рассматриваются методы и алгоритмы обработки естественных языков, проводится их сравнительный анализ. Также проводится обзор уже существующих технических решений. Затем, на основании исследований, делаются выводы и формируются требования к разрабатываемому проекту.

Далее проводится проектирование архитектуры. Выделяются основные программные компоненты, проводится их детальный обзор. Затем проектируется план развертывания системы в датацентре с обеспечением отказоустойчивости.

В конце работы описываются наиболее важные детали разработки программного средства. Обосновывается выбор архитектуры, строится структура решения, создаются и реализуются интерфейсы. Также описывается процесс разработки программного и графического интерфейсов. Подводятся итоги диссертации.

ЗАКЛЮЧЕНИЕ

В результате выполнения магистерской диссертации был проведен детальный анализ предметной области распознавания естественного языка. Изучены основные проблемы и задачи. Спроектировано и реализовано программное средство для регистрации отношения диалоговой реплики к тематической категории.

В сфере распознавания естественного языка были решены следующие проблемы:

- очистка текстовых данных от малозначимых символов;
- выделение токена текста;
- лемматизация и стемминг текстовых данных;
- определение значимости отдельных слов в тексте;
- определение количественной характеристики схожести двух текстов.

В результате решения проблем было построено микросервисное приложение на платформе ASP.NET Core, включающее в себя следующие обязательные компоненты:

- web-приложение;
- сервис лемматизации и стемминга;
- база данных;
- хранилище Redis.

Спроектирована схемы развертывания для обеспечения отказоустойчивости и распределения нагрузки. Развертывание реализовано в операционной системе Windows Server с использованием прокси-сервера IIS.

Были реализованы программный и графический интерфейсы.

Программный интерфейс спроектирован по принципам REST и реализует четыре группы методов, которые позволяют гибко управлять приложением. Наличие программного интерфейса позволяет интегрировать программный модуль с другими системами, а графическое отображение методов в Swagger сильно упрощает понимание структуры и возможностей программного интерфейса.

Графический интерфейс представляет собой одностраничное браузерное приложение, написанное с использованием библиотеки React. Одностраничное приложение позволяет работать с интерфейсом приложения без перезагрузки страницы браузера. Это увеличивает значительно скорость и удобство работы. Элементы управления расположены таким образом, чтобы минимизировать пользовательские движения и ускорить получение результата выполняемой операции.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Oracle 10g DBA I. Руководство слушателя. - М.: McGraw-Hill, 2005. - 600 с.
2. Аликина, Е.В. Переводческая семантография. Запись при устном переводе / Е.В. Аликина. - М.: АСТ, 2007. - 156 с.
3. Арбиб, М. Метафорический мозг / М. Арбиб. - Москва: Огни, 2010. - 837 с.
4. Архипенков, С. Хранилища данных. От концепции до внедрения / С. Архипенков, Д. Голубев, О. Максименко. - М.: Диалог-Мифи, 2002. - 528 с.
5. Бионические информационные системы и их практические применения / Коллектив авторов. - Москва: Машиностроение, 2011. - 325 с.
6. Вихнин, А. Г. Штурм четвертого мегапроекта. Кто будет новым Биллом Гейтсом? / А.Г. Вихнин, Н.З. Сакипов. - М.: Диалог-Мифи, 2008. - 288 с.
7. Гусева, А. И. Сети и межсетевые коммуникации. Windows 2000. Учебник / А.И. Гусева. - М.: Диалог-Мифи, 2002. - 256 с.
8. Искусственный интеллект и принятие решений, №1, 2014. - Москва: Высшая школа, 2014. - 143 с.
9. Искусственный интеллект и принятие решений, №3, 2015. - М.: ИСА РАН, 2015. - 630 с.
10. Искусственный интеллект. Десятая национальная конференция по искусственному интеллекту с м / Т.1-3. - Москва: ИЛ, 2008. - 359 с.
11. Кальп, Брайан Администрирование Windows Vista. Полное руководство / Брайан Кальп. - М.: Русская Редакция, БХВ-Петербург, 2008. - 672 с.
12. Каратыгин Access 2000 на примерах. Руководство пользователя с примерами / Каратыгин, Сергей. - М.: Лаборатория Базовых Знаний, 2000. - 376 с.
13. Карпова, Т.С. Базы данных. Модели, разработка, реализация / Т.С. Карпова. - М.: СПб: Питер, 2001. - 304 с.
14. Кенин, А.М. Практическое руководство системного администратора / А.М. Кенин. - М.: БХВ-Петербург, 2010. - 953 с.
15. Кокорева, Ольга Реестр Windows XP / Ольга Кокорева. - М.: БХВ-Петербург, 2005. - 548 с.
16. Коробко, И. В. PowerShell как средство автоматического администрирования / И.В. Коробко. - М.: ДМК Пресс, 2012. - 442 с.
17. Коробко, И.В. Справочник системного администратора по программированию Windows / И.В. Коробко. - М.: БХВ-Петербург, 2009. - 1000 с.
18. Кузнецов, С.Д. Основы баз данных / С.Д. Кузнецов. - М.: Бином, 2007. - 484 с.

19. Малыгина, М. Базы данных: основы, проектирование, использование / М. Малыгина. - М.: БХВ-Петербург, 2004. - 512 с.
20. Нортон, П. Полное руководство по Microsoft Windows XP / П. Нортон, Д. Мюллер. - М.: ДМК Пресс, 2011. - 736 с.
21. Озкарахан, Э. Машины баз данных и управление базами данных / Э. Озкарахан. - М.: Мир, 1989. - 696 с.
22. Потапова, Р.К. Тайны современного Кентавра. Речевое взаимодействие "человек-машина" / Р.К. Потапова. - Москва: СИНТЕГ, 2003. - 152 с.
23. Рамбиди, Николай Нанотехнологии и молекулярные компьютеры / Николай Рамбиди. - Москва: РГГУ, 2007. - 537 с.
24. Руссинович, М. Внутреннее устройство Microsoft Windows / М. Руссинович, Д. Соломон. - М.: Питер, 2013. - 847 с.
25. Руссинович, Марк Внутреннее устройство Microsoft Windows. Основные подсистемы ОС / Марк Руссинович. - М.: Питер, 2012. - 497 с.
26. Скачиваем фильмы, музыку и программы из Интернета. Пиринговые сети. eDonkey, BitTorrent, KaZaA, DirectConnect. - М.: Наука и техника, 2006. - 272 с.
27. Управление в условиях неопределенности. - М.: Издательство СПбГТУ, 2002. - 400 с.
28. Флах, Петер Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных. Учебник / Петер Флах. - М.: ДМК Пресс, 2015. - 400 с.
29. Шамис, А. Л. Поведение, восприятие, мышление: проблемы создания искусственного интеллекта / А.Л. Шамис. - М.: Едиториал УРСС, 2005. - 224 с.
30. системы, Открытые Открытые системы. СУБД №02/2011 / Открытые системы. - М.: Открытые Системы, 2011. - 416 с.
31. Jesse, Russell Ван Перси, Робин / Jesse Russell. - М.: VSD, 2012. - 857 с.
32. Ашманов, И.С. Идеальный поиск в Интернете глазами пользователя / И.С. Ашманов. - М.: Питер, 2011. - 168 с.
33. Веллинг, Л. Разработка веб-приложений с помощью PHP и MySQL / Л. Веллинг. - М.: Диалектика / Вильямс, 2016. - 811 с.
34. Дейв, Марк Разработка приложений для iPhone, iPad и iPod touch с использованием iOS SDK / Марк Дейв. - М.: Диалектика / Вильямс, 2012. - 697 с.
35. Загуменнов, А.П. Как раскрутить и разрекламировать Web-сайт в сети Интернет / А.П. Загуменнов. - М.: Книга по Требованию, 2005. - 384 с.
36. Ломов, Артемий Самоучитель Apache, Perl, MySQL. Практика создания динамических сайтов (+ CD-ROM) / Артемий Ломов. - М.: БХВ-Петербург, 2007.
37. Нолан, Хестер Как создать превосходный сайт в Microsoft

- Expression Web 2 и CSS / Хестер Нолан. - М.: ДМК Пресс, 2009. - 959 с.
38. Рассел, Джесси Спартак-Нальчик / Джесси Рассел. - М.: VSD, 2012. - 863 с.
39. Томлинсон, Тодд CMS Drupal 7. Руководство по разработке системы управления веб-сайтом / Тодд Томлинсон. - М.: Вильямс, 2011. - 560 с.
40. Фримен, Э. Паттерны проектирования / Э. Фримен. - М.: Питер, 2017. - 169 с.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

- 1–А. Анисько, Д.Г. Внедрение показателей КРІ / Д. Г. Аниско, П. А. Федорова, А. Г. Давыдовский //Республиканский конкурс научных работ студентов. – Минск, 2017.