УДК 004.89

ГОЛОСОВОЕ УПРАВЛЕНИЕ НА САЙТЕ: ПРИНЦИП РАБОТЫ И НЕОБХОДИМЫЕ ИНСТРУМЕНТЫ

Габитова А.Р., студент гр. 4410, IV курс
Научный руководитель: Валитова Н.Л., к.т.н., доцент
Казанский национальный исследовательский технический университет им. А.Н. Туполева – КАИ
г. Казань

Статья посвящена теме голосового управления на сайте. В работе изучается понятие голосового управления, общие принципы обработки текста в речь и наоборот. Приводится обзор сервисов, при помощи которых можно подключить голосовую обработку на сайте. А также рассматриваются плюсы и минусы использования голосового управления.

В современном мире все больше людей используют голосовые помощники в повседневных делах: позвонить другу, заказать такси, включить музыку, запустить робот-пылесос и т.д. Для многих людей использование такого рода технологий является явным удобством, ведь подобные команды произнести легко и быстро. Плюсы очевидны и видны.

Голосовые помощники построены на технологиях распознавания голоса и машинного обучения. Принцип работы голосового помощника довольно прост: необходимо считать звук, отфильтровать сигнал, «перевести» его в цифровой код, проанализировать и выполнить шаблонную команду, если она присутствует в шаблонах голосового помощника. Голосовые ассистенты являются неотъемлемой частью нашей жизни и применяются повсеместно. Именно поэтому вопрос подключения и реализации голосового управления на сайте является актуальным [1].

Поставим целью – изучить реализацию голосового управления на сайте. Для этого необходимо понять:

- что понимается под «голосовым управлением»;
- как реализуется обработка текста в речь и речи в текст;
- как можно подключить голосовое управление на сайте, какие инструменты необходимы для этого;
- плюсы и минусы в использовании голосового управления на сайте.

Что понимается под голосовым управлением

Голосовое управление - возможность управлять устройством с помощью голосовых команд, без нажатия кнопок. Чтобы активировать какую-либо команду, достаточно проговорить ее название. Оно идеально подходит для многих людей, которые не могут управлять компьютером при помощи рук [2].

Голосовое управление основано на распознавании речи. Основная задача такого элемента управления – распознать в речи пользователя те команды, которые записаны в программе и выполнить соответствующие им действия.

Например, необходимо перелистнуть страницу сайта на следующую и известно, что для это используется команда «дальше», тогда для перехода к следующей странице необходимо четко произнести команду «дальше».

В понятие «голосовое управление» не редко включается и возможность прослушать ответ устройства на команду. Такая реакция упрощает восприятие человеком информации, полученной при работе с тем или иным приложением.

Таким образом, голосовое управление – это возможность взаимодействовать с устройством при помощи голоса, а именно прослушать информацию и выполнять действия при помощи голосовых команд.

Принцип обработки речи в текст

Для того, чтобы использовать механизм, необходимо понимать, как он устроен. Именно поэтому стоит разобраться, как именно происходит обработка и преобразование текста в речь и наоборот.

Обработка голоса включает несколько этапов, от записи звука до выполнения команды.

Этап 1: Получение звукового сигнала

Прежде чем обрабатывать звук, необходимо его получить. Микрофон пользователя воспринимает речь и записывает ее в виде звукового сигнала, который отправляется на сервер. Там убираются лишние шумы и помехи, сигнал разбивается на фонемы – фрагменты длительностью до 25мс. На этом этапе важно обеспечить качественный звук и минимизировать фоновые шумы [3].

Этап 2: Преобразование речи в текст (ASR)

С помощью алгоритмов Automatic Speech Recognition (ASR) аудио преобразуется в текст. В основе технологии используется принцип соотношения звука и слова при помощи искусственного интеллекта. На данном этапе специально обученная нейросеть определяет по спектрограмме аудиозаписи, какой букве соответствует тот или иной рисунок сигнала. Затем ее задача – преобразовать отдельные буквы в слова, а слова – в предложения.

Для того, чтобы нейронная сеть могла распознавать буквы и слова, используется подготовленный набор данных в формате «аудио-текст». Нейросеть в процессе обучения учится определять вероятность того, какая буква перед ней по рисунку аудиодорожки. Далее нейронная сеть собирает буквы в слова, подбирая их из словаря (например, словарь определенных терминов или некоторого языка) [4].

Помимо нейронных сетей используются и другие методы анализа речевых сообщений: анализ с использованием преобразования Фурье, анализ с использованием вейвлет преобразования, кепстральный анализ, анализ с использованием линейного предсказания и другие [5].

Этап 3: Анализ текста (NLP)

Получив некоторый набор слов, искусственный интеллект определяет правильное расположение их в тексте, а также расставляет знаки препинания. Для обработки набора слов и преобразования их в связный текст используется Natural Language Processing или Обработка естественного языка. «Обработка естественного языка (Natural Language Processing, NLP) — пересечение

машинного обучения и математической лингвистики, направленное на изучение методов анализа и синтеза естественного языка» [6].

Для анализа текста используются различные подходы. В основе таких подходов лежит тот или иной способ разбиения слов на группы. Вот определения некоторых из них:

- токенизация (иногда – сегментация) по предложениям – процесс разделения письменного языка на предложения-компоненты;
- токенизация (иногда – сегментация) по словам – процесс разделения предложений на слова-компоненты;
- стемминг – грубый эвристический процесс, который отрезает «лишнее» от корня слова (часто это приводит к потере словообразовательных суффиксов);
- лемматизация – более тонкий процесс, который использует словарь и морфологический анализ, чтобы в итоге привести слово к его лемме– нормальной (словарной) форме [7].

Таким образом, задачей NLP является обработка и преобразование слов к привычным для восприятия человеком формам.

Этап 4: Выполнение команды

Когда речь преобразована в текст и программист имеет перед собой некоторый набор данных от пользователя, он может использовать его для выделения определенных команд для выполнения. Если пользователем была произнесена некоторая команда, то на сайте выполняется соответствующее действие: открывается нужный раздел, показываются результаты поиска, проговаривается текст и т.д.

Принцип обработки текста в речь

Text-to-Speech (TTS) — технология синтеза речи, которая преобразует текст в аудио. Это может быть важно для: озвучивания ответов на запросы (например, подтверждение заказа), помощи пользователям с нарушениями зрения и других подобных задач.

Text-to-Speech или синтез речи использует тот же принцип, что и технологии распознавания речи, но наоборот. Сначала обрабатывается текст, затем он преобразуется в речь. Для преобразования текста в речь используются нейронные сети, сопоставляющие звук и определенные буквы или слова, используются специальные подходы в обработке текстов для их анализа и преобразования [8].

Из-за попыток разработчиков добиться человекоподобного звучания синтез речи технологически сложнее, чем распознавание речи.

Процесс синтеза речи также можно разделить на этапы.

Этап 1: Нормализация текста

На данном этапе текст приводится в удобный для обработки формат – текстовый. Производится замена чисел, дат в текстовый формат, расшифровка аббревиатур, сокращений и так далее. Иными словами, все значения, которые сокращают запись нашей речи, переводятся обратно в разговорный текст.

Этап 2: Лингвистический анализ

Выделим в этап «лингвистический анализ» все, что связано с формированием правильного произношения текста в соответствии с его смыслом. Тогда лингвистический анализ будет включать в себя некоторые следующие подзадачи: выявление и понимание омографов, проставление ударений в соответствии с правильным смыслом слова, разбиение текста по смыслу и проставление пауз. Для повышения «человечности» в произношении необходимо также проставить смысловые ударения и выделить аллофоны [8].

Этап 3: Воспроизведение звука

На этом этапе полностью обработанный тест сопоставляют со звуком. Для этого используется нейронная сеть, обучение которой очень схоже с обучением сети для распознавания речи. Для ее обучения используются наборы данных в формате «текст-аудио». Нейросеть учится устанавливать соответствие рисунка сигнала аудиозаписи с текстом.

Инструменты для внедрения голосового управления на сайте

Рассмотрим, как же можно подключить голосовое управление на сайте. Для интеграции можно использовать готовые API и сервисы.

1. Web Speech API

Web Speech API – это нативный API браузеров (поддерживается Edge, Chrome). Может использоваться как для распознавания речи через микрофон, так и для синтеза речи при озвучивании текстов.

2. Yandex SpeechKit

Yandex SpeechKit — это облачная платформа для распознавания речи (ASR) и синтеза текста в речь (TTS). Она поддерживает несколько языков, включая русский, и предлагает высокое качество обработки голоса благодаря нейросетевым моделям. SpeechKit легко интегрируется через API и подходит для создания голосовых помощников, чат-ботов и других приложений с голосовым управлением.

3. Google Dialogflow

Google Dialogflow – это платформа для разработки голосовых интерфейсов и чат-ботов. Платформа может обеспечить NLP для анализа запросов и интеграцию с Google Assistant.

4. Microsoft Azure Cognitive Services

Microsoft Azure Cognitive Services представляет собой набор инструментов для распознавания и синтеза речи. Из особенностей можно выделить поддержку 120+ языков и возможность адаптировать модели под специфичные термины (например, медицинские)

5. Alan AI

Alan AI — это платформа для добавления голосового управления в приложения и сайты. Она позволяет создавать голосовых ассистентов с поддержкой естественного языка (NLP) и легко интегрируется через JavaScript. Alan AI подходит для навигации, управления контентом и автоматизации задач, предлагая простую настройку и поддержку мультиязычности.

Плюсы и минусы голосовой адаптации

Причины, по которым стоит использовать голосовое управление на сайте:

1. Доступность. Интеграция голосового управления на сайте важна, в первую очередь, для людей с нарушением зрения. Более 2,2 млрд людей в мире имеют нарушения зрения. Для них экранные дикторы (например, JAWS, NVDA) и голосовые команды становятся основным инструментом навигации [9]. Кроме того, ГОСТ Р 52872-2019 и WCAG 3.0 требуют обеспечения альтернативных методов взаимодействия, включая голосовой ввод [10];

2. Удобство. Позволяет эффективно пользоваться сайтом в условиях, когда у человека заняты руки или нет возможности смотреть на экран (например, во время вождения);

3. Обеспечивает человеку мультизадачность. Пользователи могут параллельно выполнять другие задачи;

4. Инновационный имидж. Внедрение голосового управления – показатель технологической осведомленности и несомненный плюс для компании.

Однако есть и недостатки подобной адаптации на сайте.

1. Точность распознавания. Фоновый шум, акценты, омонимы и прочие внешние воздействия ухудшают точность распознавания, что может привести к неправильным исполнениям команд;

2. Ограниченность. Не все функции, доступные на сайте, возможно выполнить при помощи голосовых команд. Так, например, пароль вводить при помощи клавиатуры будет удобнее, то же самое касается и любых текстов со специфичными знаками;

3. Конфиденциальность. Вид постоянно включенного микрофона так или иначе заставляет задуматься, где хранятся данные после обработки и хранятся ли вообще;

4. Технические требования. Зависимость от скорости интернета и поддержки браузеров.

Заключение

Голосовое управление — не тренд будущего, а реальность, которая уже меняет веб-интерфейсы. Именно поэтому, необходимо изучать способы реализации технологии на сайтах.

Для внедрения голосового управления можно выбрать один из двух подходов: попытаться обучить нейронную сеть самостоятельно или подключить один из готовых сервисов. Первый способ требует не малых знаний и усилий от разработчика. Что касается второго способа, одним из основных инструментов для реализации голосового управления в веб-приложениях остается Web Speech API. Он подходит для базовых задач, является бесплатным и легко подключается – достаточно навыков работы с JavaScript.

Список литературы:

1. Поначугин А.В., Пичужкина Д.Ю., Смекалова Е.С. – Голосовой помощник как технология обработки данных [Электронный ресурс] / Наука без границ: электрон. научн. журн. 2020. Режим доступа:

https://cyberleninka.ru/article/n/golosovoy-pomoschnik-kak-tehnologiya-obrabotki-dannyh/viewer (дата обращения: 17.03.2025)

2. Речевые технологии. Голосовое управление / [Электронный ресурс] // : [сайт]. – URL: https://speetech.by/technologies/golosovoe-upravlenie#fragment-2/ (дата обращения: 15.03.2025)

3. Технология распознавания речи и ее значение для бизнеса / [Электронный ресурс] // : [сайт]. – URL: Технология распознавания речи: что это такое, как работает, где применяется и какие бизнес задачи решает | Блог MWS (дата обращения: 16.03.2025)

4. Как работает распознавание речи / [Электронный ресурс] // : [сайт]. – URL: https://developers.sber.ru/help/salutespeech/how-speech-recognition-works (дата обращения: 16.03.2025)

5. Муратов Н.А. – Основные методы обработки речевых сообщений [Электронный ресурс] / Новые информационные технологии в автоматизированных системах: электрон. научн. журн. 2018. Режим доступа: https://cyberleninka.ru/article/n/osnovnye-metody-obrabotki-rechevyh-soobscheniy/viewer (дата обращения: 17.03.2025)

6. Обработка естественного языка / [Электронный ресурс] // : [сайт]. – URL: https://neerc.ifmo.ru/wiki/index.php?title=%D0%9E%D0%B1%D1%80%D0%B0%D0%B1%D0%BE%D1%82%D0%BA%D0%B0_%D0%B5%D1%81%D1%82%D0%B5%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D0%BE%D0%B3%D0%BE_%D1%8F%D0%B7%D1%8B%D0%BA%D0%B0 (дата обращения: 16.03.2025)\

7. Основы Natural Language Processing для текста / [Электронный ресурс] // : [сайт]. – URL: https://habr.com/ru/companies/Voximplant/articles/446738/ (дата обращения: 16.03.2025)

8. Что такое синтез речи / [Электронный ресурс] // : [сайт]. – URL: https://developers.sber.ru/help/salutespeech/creating-audio-from-text (дата обращения: 16.03.2025)

9. Как верстальщику адаптировать сайт для людей с ограниченными возможностями / [Электронный ресурс] // : [сайт]. – URL: https://wim.agency/blog/article89-kak-verstalshchiku-adaptirovat-sajt-dlya-lyudej-s-ogranichennymi-vozmozhnostyami/ (дата обращения: 16.03.2025)

10. ГОСТ Р 52872-2012 Интернет-ресурсы. Требования доступности для инвалидов по зрению / [Электронный ресурс] // : [сайт]. – URL: https://slabovid.ru/info/requirements2012/ (дата обращения: 17.03.2025)