

УДК 004.8

ПРИНЦИП РАБОТЫ ГОЛОСОВОГО ПОМОЩНИКА ЯНДЕКС АЛИСА
THE PRINCIPLE OF OPERATION
OF THE YANDEX ALICE VOICE ASSISTANT

К.С. Кухновец, В.В. Суша, М.М. Ровдо

Научный руководитель – Г.А. Михальцевич, старший преподаватель
Белорусский национальный технический университет, г. Минск

K. Kuhnovets, V. Susha, M. Rovdo

Supervisor – G. Mikhaltsevich, Senior Lecturer
Belarusian national technical university, Minsk

Аннотация: В наше время все чаще используют все сторонние интернет помощники. Один из таких помощников “Яндекс Алиса” и будет рассматриваться в данной работе. Речь пойдет о проявлениях и функциях искусственного интеллекта в современных голосовых помощниках.

Abstract: Nowadays, all third-party Internet assistants are increasingly used. One of these assistants is “Yandex Alice” and will be considered in this paper. We will talk about the manifestations and functions of artificial intelligence in modern voice assistants.

Ключевые слова: голосовой помощник, Яндекс Алиса, искусственный интеллект.

Key words: voice assistant, Yandex alice, Yandex alice, artificial intelligence.

Введение

Яндекс Алиса – это голосовой помощник, который может общаться с пользователем на разные темы, помогать в поиске информации, управлять умным домом, включать музыку и многое другое. Распознаёт естественную речь, имитирует живой диалог, даёт ответы на вопросы пользователя и благодаря запрограммированным навыкам решает прикладные и трудные задачи. Алиса работает на смартфонах, планшетах, компьютерах и в навигаторах автомобилей. По данным Яндекса месячная аудитория Алисы по состоянию на декабрь 2019 года составила 45 млн человек при более миллиарде запросов за год. Подобные сервисы, такие как Siri, Google Assistant, Alexa и Cortana, также используют искусственный интеллект для распознавания и обработки голосовых запросов. В этих тезисах будет изложено, как работает Яндекс Алиса и другие голосовые помощники.

Основная часть

В данной работе речь пойдет про искусственный интеллект и о его проявлениях в современном мире. Разберемся в том, как же так получилось, что в наших, например, карманных сотовых телефонах, есть нечто, что не только способно понимать нашу речь, но и может выполнять некоторые команды пользователя. Если не углубляться в философскую составляющую вопроса, и не копать до трудов философов нового времени, искусственный интеллект берет начало с появления первых вычислительных машин в середине двадцатого века. Это были устройства, которые являлись элементарными калькуляторами,

правда, имевшие тогда большие размеры. По мере развития самих компьютеров, которые постепенно становились все компактнее, совершенствовались и алгоритмы, благодаря которым они работают и в наше время. Вначале эпохи вычислительных машин сложить два и два и получить 4, было огромным достижением. Сейчас мы приблизились к порогу создания, может и не до конца развитых, но весьма способных программ, основанных на искусственных нейросетях, которые уже умеют самообучаться и не только обыгрывать гроссмейстеров в шахматы, но даже сочинять музыку. Все это, конечно же, не могло не отразиться на индустрии пользовательских гаджетов и возможности смартфонов стали такими, как в настоящее время. Начали появляться и голосовой помощник, первым же приложением – голосовым помощником, как ни странно, было *siri*, разработка которой велась международным центром искусственного интеллекта с 2007 года, да это было именно приложение, которое распространялось через *skype*. Пользоваться им могли только владельцы айфонов. Фирма *Apple* быстро положила глаз на перспективную разработку, и в апреле 2010 года, выкупила ее. Следующая зашевелилась компания *amazon* и примерно с 2010 года стала разрабатывать собственный аналог *siri* и уже в 2014 году выпустила собственную фирменную «алексу». Распространялась она, как и *siri*, в виде приложения. Но к ней добавилась еще поддержка *android*. Но на этом разработчики не остановились и видимо поняли, что высокой прибыли с приложения они не соберут, приняли решение сделать отдельное устройство, которое будет работать исключительно как голосовой помощник. Так на свет появилась первая умная колонка с виртуальным помощником в ней. *Amazon Echo* – это было уже умное домашнее устройство. Вроде умных термостатов и умных лампочек. А именно такого центрального мозга не хватало всей системе. Уже сейчас помимо стандартных запросов от пользователя, *amazon* умеет управлять умными домашними приборами удаленно, что со временем, конечно, научились делать и помощники других конкурирующих компаний. Совсем недавно одна отечественная компания выпустила свой собственный голосовой помощник, речь идет о продукте под названием Алиса. Основным преимуществом отечественного голосового помощника, безусловно, является то, что оно способно поддерживать какой-никакой диалог и понимать контекст в этом диалоге, например – по одному и тому же запросу она может перечислять несколько вариантов ответа. У Алисы есть определенное количество запросов, на которые она может выдавать ответы сразу. Как ни странно, но число это довольно ограничено. Если понимание текста и поддержание разговора для лиц в принципе не составляет особого труда, то в остальных некоторых аспектах она всё ещё не настолько умна. Например, у нее есть проблемы с пониманием речи в некоторых моментах, то есть от раза к разу Алиса может неправильно понимать ваши слова, если вы используете какие-нибудь просторечия, то она вас тоже, скорее всего, не поймет. Стоит выделить ряд достоинств: она понимает контекст, у нее очень хорошая русская озвучка. Русская, но естественно, для русскоязычной аудитории. В итоге можно сказать, что искусственный интеллект и голосовые помощники в наших гаджетах – это все еще дело будущего, несмотря на то, что они уже много чего умеют и могут вести элементарный

диалог с пользователем, но до полной интеграции с окружающим миром, всему этому, еще далеко.

Распознать голосовой запрос Алисе помогает технология SpeechKit. На этом этапе происходит отделение голоса от фоновых шумов. Разобраться с аспектами, диалектами, сленгами и англицизмами алгоритмам позволяет накопленная Яндексом база из миллиарда, произнесённых в разных условиях, фраз.

На следующем этапе надо наделить запрос смыслом и подобрать правильный ответ. Технология Turing, своим названием отсылает к Алану Тьюрингу и его тесту. Благодаря ей Алиса может не только давать ответы на конкретные вопросы, но и общаться с пользователем на отвлечённые темы. Для этого текст запроса дробится на отдельные слова, которые в дальнейшем отдельно анализируются. Для максимально точного ответа Алиса учитывает историю взаимодействия с ней, интонацию запроса, предыдущие фразы. Это объясняет тот факт, что на один вопрос разные пользователи могут получить разные ответы.

Первоначально нейросеть Алиса обучалась на массивах текстов из русской классики, включая произведения Льва Толстого, Фёдора Достоевского, Николая Гоголя и других. Затем Алиса обучалась на массивах живых текстов из Интернета. Как рассказывал глава управления машинного обучения Яндекса Михаил Биленко, в интервью изданию Meduza, в ходе ранних тестов в манере общения Алисы появилась дерзость, которая удивляла и забавляла пользователей. Чтобы дерзость не перешла в грубость и чтобы ограничить рассуждения Алисы на темы, связанные с насилием, ненавистью или политикой, в голосового помощника была внедрена система фильтров и стоп-слов.

Последний этап – озвучивание ответа – реализуется с помощью технологии Text-to-speech. Основой служат записанные в студии 260 тысяч слов и фраз, которые затем были «порезаны» на фонемы. Из этой базы нейросеть собирает ответ, после чего интонационные перепады в готовой фразе сглаживаются нейросетью, что приближает речь Алисы к человеческой. При этом нужно отличать голосовых помощников, как Алиса и Маруся, от роботов для телефонов, как Маша и прочие.

Для того, чтобы понять и ответить на голосовой запрос, помощнику нужно выполнить несколько этапов:

- Распознавание речи. На этом этапе помощник преобразует звуковой сигнал в текстовый вид. Для этого он делит запись голоса на короткие фрагменты, которые анализируются с помощью математических преобразований и нейронных сетей. Помощник определяет фонемы — элементарные звуковые единицы языка, и на основе вероятностной модели и контекста фразы составляет варианты слов, которые могли быть сказаны.
- Понимание намерений. На этом этапе помощник анализирует текстовый запрос, чтобы выявить цель пользователя. Для этого он использует методы обработки естественного языка (NLP), такие как синтаксический и семантический анализ, классификацию и извлечение сущностей.

Помощник определяет тему запроса, ключевые слова, параметры и дополнительную информацию, которая нужна для выполнения задачи.

- Генерация ответа. На этом этапе помощник формирует ответ на запрос, который может быть текстовым, голосовым или графическим. Для этого он может обращаться к разным источникам данных, таким как поисковые системы, базы знаний, приложения или сервисы. Помощник также может задавать уточняющие вопросы, если ему не хватает информации для выполнения запроса. Для генерации голосового ответа помощник использует технологию синтеза речи, который преобразуют текстовую информацию в звук, с учетом интонации, акцента и эмоций.

Яндекс Алиса работает на основе нейронных сетей и глубокого обучения (Deep Learning). Именно за счет этого она способна самостоятельно обучаться, а также воспринимать не только заготовленные команды, но и произвольные фразы.

Заключение

Яндекс Алиса зарекомендовала себя как отличный голосовой помощник, благодаря своей функциональности, семантикой, простым интерфейсом и словарным запасом. С Алисой уже успешно работают, в странах СНГ. А также, с некоторыми ограничениями, в Бельгии, Германии, Израиле, Испании, Латвии, Нидерландах, США, Турции, Финляндии, Чехии, Эстонии и в других странах.

Литература

1. Ушкин, С.Г. Жить с Алисой: как голосовые помощники трансформируют коммуникации / С.Г. Ушкин, Е.А. Коваль, А.Н. Яськин // Сборник: Искусственный интеллект, факторы неравенства и социальная политика. Том 20, №3 – Москва, 2022. – С. 361-376.
2. Wikipedia [Электронный ресурс]. – Электронные данные. – Режим доступа: <https://ru.wikipedia.org/>. – Дата доступа: 02.10.2023.