

Virtual Personal Assistant Apps development

Problem Statement

Artificial Intelligence personal assistants have become plentiful over the last few years. Applications such as Siri, Bixby, Ok Google and Cortana make mobile device users' daily routines that much easier. You may be asking yourself how these functions. Well, the assistants receive external data (such as movement, voice, light, GPS readings, visually defined markers, etc.) via the hardware's sensors for further processing - and take it from there to function accordingly.

Not too long ago, building an AI assistant was a small component of developers' capacities; however, nowadays, it is quite a realistic objective even for novice programmers. To create a simple personal AI assistant, one simply needs dedicated software and around an hour of working time. It would take much more time, though, to create something more advanced and conceptually innovative. Nonetheless, well thought-out concepts can result in a great base for a profitable startup. Let us consider the six most renowned applications based on artificial intelligence concepts that can help create your virtual AI assistant app.

Background

Siri. Siri is Apple Inc.'s cloud software that can answer users' various questions and give recommendations, due to its voice processing mechanisms. When in use, Siri studies the user's preferences (like contextual advertising) to provide each person with an entirely individual approach. This software solution is also useful for developers; the presence of API called SiriKit provides smooth integration with new applications developed for iOS and watchOS platforms.

Ok Google. Ok Google is an Android-based voice recognition application, which is launched by users uttering commands of the same name. This software features very advanced functions including web search, route optimization, memo scheduling etc. that can collectively help users solve a wide array of daily tasks. Like Siri, the creators of Ok Google offer Google Voice Interaction API. This interface can become a truly indispensable tool in the development of mobile applications for the Android platform.

Cortana. A virtual intelligent assistant with the function of voice recognition and AI elements, Cortana was developed for such platforms as Windows, iOS, Android, and Xbox One. It can predict users' wants and needs based on their search requests, e-mails, etc. One of Cortana's distinguishable features is her sense of humor. "She" can sing, make jokes and speak to users informally.

Amazon Echo. Amazon Echo combines hardware and software that can search the web, help with scheduling of upcoming tasks and play various sound files all based on voice recognition. A small speaker equipped with sound sensors, the device can be automatically activated by exclaiming "Alex."

Nina. Software with AI elements that has a main goal of narrowing down the amount of physical effort spent on the solution of daily tasks (web search, scheduling, etc.) Due to elaborate analytical mechanisms, Nina becomes "smarter" with every day of personal utilization.

Bixby. Samsung’s Bixby application is another successful implementation of the AI concept. It also builds a unique user approach, based on interests and habits. Bixby features advanced voice recognition mechanisms, and uses the camera to identify images, based on markers and GPS.

Methodology

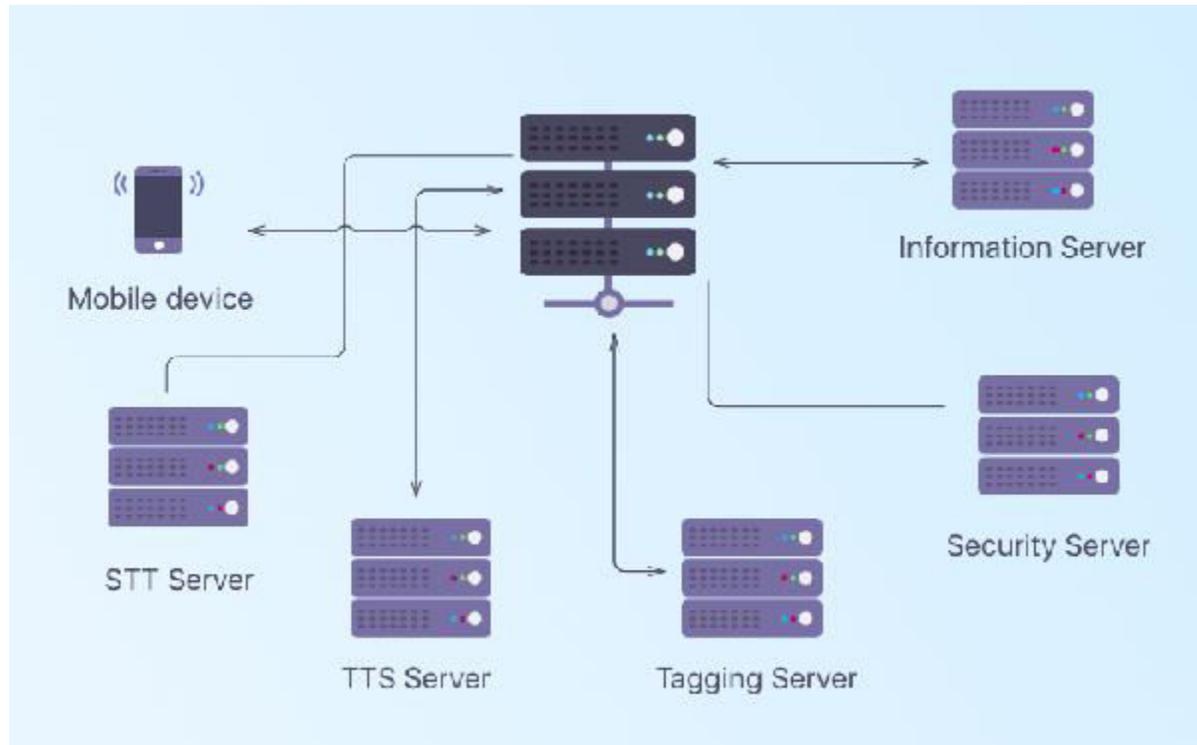


Fig: 1 Mobile voice assistant’s architecture

The general operating principle of artificial intelligence assistants is the ability to make personal decisions based on incoming data. The software has to include an advanced set of tools for processing received data, in order to make proper individual choices. Artificial neural networks were invented to help develop the discussed software. Such networks imitate the human brain’s ability to remember, to help the assistant recognize and classify data and customize predicting mechanisms based on thorough analysis. The memory process is executed deductively, i.e. top-down: first, the app analyzes several variants of outcome; then, it remembers the variants applied by a human (i.e. the system remembers proper answers to the question “How are you?” such as “I’m fine”, “Not very well” etc., and ignores answers like “Yes”, “No” and others) and “self-educates” to be able to generate situation-based algorithms later. It is not necessary to manually enter information into the app to build your own personal artificial intelligence assistant. API software was developed for that, and the application programming interface aids the apps in the recognition of faces, speech, documents and other external factors. There are a number of APIs on the market, most popular of which are api.ai, Wit.ai, Melissa, Clarifai, Tensorflow, Amazon AI, IBM Watson; with less widespread options including Cogito, DataSift, iSpeech, Microsoft Project Oxford, Mozscape and OpenCalais. Let us examine some of these.

Experimental Design

How to Create Virtual Assistant Apps like Siri and Google Assistant

Developing your own voice assistant app

If you intend to make your own Siri or Google assistant, make sure that you do possess the appropriate skills and sources, because this process is far from simple. Basic technologies in mobile voice assistants.

Voice/speech to text (STT)

This is the process of converting speech signal into digital data (e.g., text data). The voice may come as a file or a stream. You can use CMU Sphinx for its processing.

Text to speech (TTS)

This is the opposite process that translates text / images in a human speech. It is very useful when, for instance, a user wants to hear the correct pronunciation of a foreign word.

Intelligent tagging and decision making

Intelligent tagging and decision making serve for interpreting the user's request. For example, the user may ask: 'What do I watch tonight?'. The technology will tag the top-rated movies and suggest you a few according to your interests. The AlchemyAPI may help you in the implementation of this task.

Image recognition

Image recognition is an optional but very useful picture. Later, you can use it for developing multimodal speech recognition. Have a look at OpenCV if you are thinking of developing it.

Noise control

The noises from cars, electrical appliances, other people talking near you make the user's voice unclear. This technology will reduce or eliminate the background noise that prevents a correct voice recognition.

Voice Biometrics

This is a very important option from the point of view of security. Thanks to this feature, the voice assistant may identify who is talking and whether it is necessary to respond. Thus, you may avoid a comic situation that happened to Siri and Amazon Alexa when they lowered the temperature in a house and even turned off someone's thermostat by hearing a relevant command from the TV speakers.

Speech compression

With this mechanism, the client side of the applications will resize the voice data and send it to the server in a succinct format. It will provide a fast application performance without annoying delays.

Voice interface

Voice interface is what the user hears and sees in return to his or her request. For the voice part, you will need to pick up the voice itself, set the rate of speech, the manner of speaking, etc. For the visual part, you will have to decide on the visual representation that a user is going to see on the screen. If reasonable, you can skip it at all. Note that voice and text data may be processed either on a server or directly within a device. In the picture below, we have shown the scheme that works with the server participation.

References

1. https://www.researchgate.net/publication/264001644_Virtual_Personal_Assistant
2. <https://artjoker.net/blog/how-to-use-artificial-intelligence-in-mobile-apps/>
3. <https://www.brainasoft.com/brainasoft/>