PERSONAL VIRTUAL ASSISTANT

Ursut Sorin Ioan

Romania

Slovakia-Zvolen-11-13 September 2024

Introduction

The Radio Rex was the first toy with voice, released in 1911. Another early device that enabled digital speech recognition was the IBM Shoebox, shown to the general public at the 1962 Seattle World's Fair after its initial release of the market in 1961.

This early computer, developed almost 20 years before the introduction of the first IBM personal computer in 1981, was able to recognize 16 spoken words and the digits 0 - 9

An intelligent virtual assistant (AVI) or intelligent personal assistant (IPA) is a software agent that can perform tasks or services for an individual based on commands or questions. Sometimes the term "chatbot" is used to refer to virtual assistants, generally or specifically accessed via online chat. In some cases, online chat programs are for entertainment purposes only. Some virtual assistants are able to interpret human speech and respond through synthesized voices. Users can ask their assistant questions, control home automation devices and media playback via voice, and manage other basic tasks like email, to-do lists, and calendars with voice commands.

As of 2017, the capabilities and use of virtual assistants are expanding rapidly, with new products entering the market and a strong focus on voice user interfaces. Apple and Google have large user bases installed on smartphones. Microsoft has a large base of Windows-based PCs, smartphones, and smart speakers.

Method of interaction with the virtual assistant

- Text (online chat), especially in an instant messaging or other application.- Voice, for example with Amazon Alexa on the Amazon Echo device, Siri on an iPhone or Google Assistant on Google-enabled / Android mobile devices.
- Taking and/or uploading images, as with Samsung Bixby on the Samsung Galaxy S

Types of virtual assistants
There are two types of virtual assistants
1. Smart advisers subject-oriented. The authors define a smart adviser as representing the concept of using automation, and digital techniques to significantly improve the consultant-client relationship.
2. Virtual assistants task-oriented. A virtual assistant is a software agent that is used to perform tasks or services for an individual.

Examples of Intelligent Virtual Assistants
Cortana, is an intelligent personal agent, figure1, made by Microsoft, and introduced in Windows 10, and Windows phones.
The name comes from the game Halo, where Cortana is a synthetically intelligent character.

Cortana

Cortana was able to set reminders, recognize natural voice without the requirement for keyboard input, and answer questions using information from the Bing search engine. Searches using Windows 10 are made only with the Microsoft Bing search engine, and all links will open with Microsoft Edge,[except when a screen reader such as Narrator was being used, where the links will open in Internet Explorer. Windows Phone 8.1's universal Bing SmartSearch features were incorporated into Cortana, which replaced the previous Bing Search app, which was activated when a user presses the "Search" button on their device. Cortana includes a music recognition service. Cortana was able to simulate rolling dice and flipping a coin. Cortana's "Concert Watch" monitored Bing searches to determine the bands or musicians that interest the user. It integrates with theMicrosoft Band watch band for Windows Phone devices if connected via Bluetooth, it was able to make reminders and phone notifications.

Since the Lumia Denim mobile phone series, launched in October 2014, active listening was added to Cortana enabling it to be invoked with the phrase: "Hey Cortana". It was able to then be controlled as usual. Some devices from the United Kingdom by O2 received the Lumia Denim update without the feature, but this was later clarified as a bug and Microsoft has since fixed it.

Cortana integrated with services such as Foursquare to provide restaurant and local attraction recommendations and LIFX to control smart light bulbs

Figure 1. Cortana Inteligence
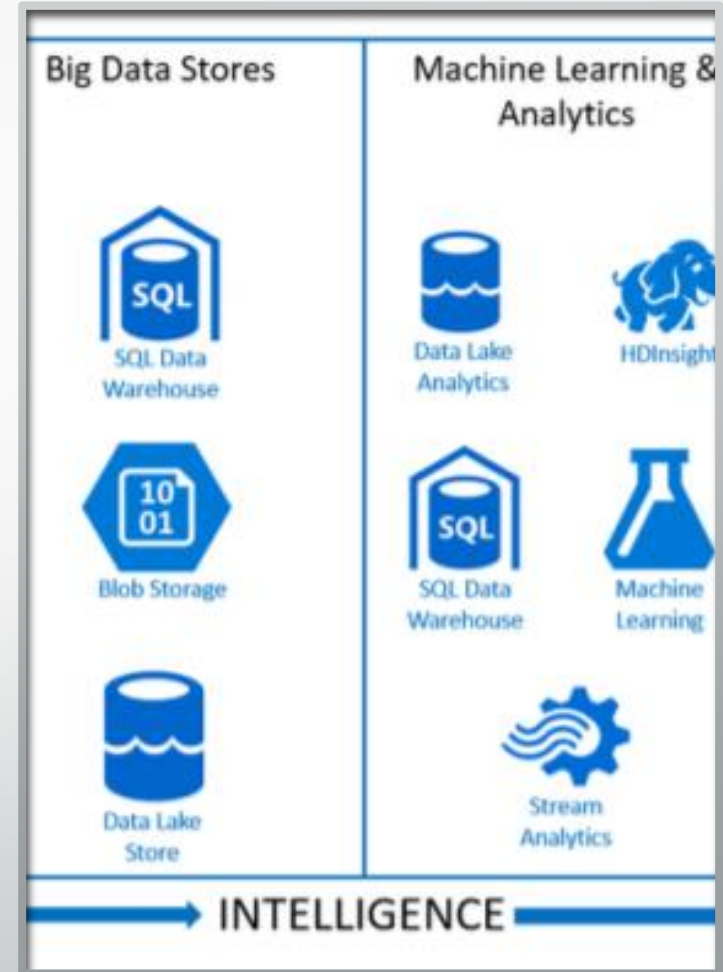
Alexa

Alexa, is an intelligent personal agent created by the Amazon Company, figure 2, which presents a multitude of functions. Unlike Siri or Cortana, Alexa is a separate device. The agent can be activated using a wait word, i.e. Echo.

Alexa can perform a number of preset functions out-of-the-box such as set timers, share the current weather, create lists, access Wikipedia articles, and many more things.Users say a designated "wake word" (the default is simply "Alexa") to alert an Alexa-enabled device of an ensuing function command. Alexa listens for the command and performs the appropriate function, or skill, to answer a question or command. When questions are asked, Alexa converts sound waves into text which allows it to gather information from various sources.

Behind the scenes, the data gathered is then sometimes passed to a variety of suppliers including WolframAlpha, iMDB, AccuWeather, Yelp, Wikipedia, and others to generate suitable and accurate answers.[33] Alexa-supported devices can stream music from the owner's Amazon Music accounts and have built-in support for Pandora and Spotify accounts.

Alexa can play music from streaming services such as Apple Music and Google Play Music from a phone or tablet.

Figure 2. The flow of executing a command for the Alexa Assistant
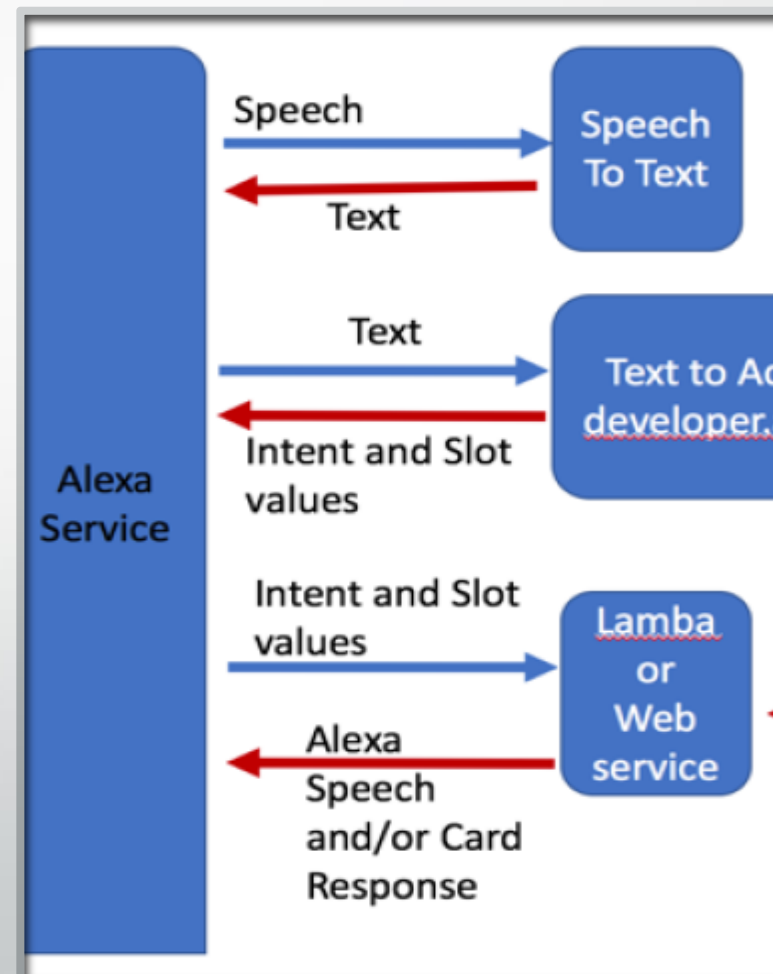
Alexa

In addition to performing pre-set functions, Alexa can also perform additional functions through third-party skills that users can enable. Some of the most popular Alexa skills in 2018 included "Question of the Day" and "National Geographic Geo Quiz" for trivia; "TuneIn Live" to listen to live sporting events and news stations; "Big Sky" for hyper-local weather updates; "Sleep and Relaxation Sounds" for listening to calming sounds; "Sesame Street" for children's entertainment; and "Fitbit" for Fitbit users who want to check in on their health stats. In 2019, Apple, Google, Amazon, and Zigbee Alliance announced a partnership to make their smart home products work together.

Amazon is enhancing Alexa with generative AI features using its Titan model, aiming to compete with AI like ChatGPT. The upgrade will be offered as a separate subscription service, potentially costing between $10 and $20 per month.

There are also humour related voice commands. One example is if you ask "Alexa, do you know GLaDOS?", Alexa will reply with "We don't really talk after what happened". This is a nod to the Portal video game franchise.

Music

   Alexa supports many subscription-based and free streaming services on Amazon devices. These streaming services include Prime Music, Amazon Music, Amazon Music Unlimited, Apple Music, TuneIn, iHeartRadio, Audible, Pandora, and Spotify Premium. However, some of these music services are not available on other Alexa-enabled products that are manufactured by companies external of its services. This unavailability also includes Amazon's own Fire TV devices or tablets.

   Alexa is able to stream media and music directly. To do this, Alexa's device should be linked to the Amazon account, which enables access to one's Amazon Music library, in addition to any audiobooks available in one's Audible library. Amazon Prime members have an additional ability to access stations, playlists, and over two million songs free of charge. Amazon Music Unlimited subscribers also have access to a list of millions of songs.

   Amazon Music for PC allows one to play personal music from Google Play, iTunes, and others on an Alexa device. This can be done by uploading one's collection to My Music on Amazon from a computer. Up to 250 songs can be uploaded free of charge. Once this is done, Alexa can play this music and control playback through voice command options.

Siri

- **Siri**, is the intelligent assistant now present in Apple products, fig.3. It appeared on October 4, 2011 with the release of the operating system, IOS5 for Iphone4S. It also supports a wide range of user commands, including making phone calls, checking basic information, scheduling events and reminders, managing device settings, searching the Internet, navigating areas, finding entertainment information, and can engage with integrated iOS apps.

- With the release of iOS 10 in 2016, Apple opened limited third-party access to Siri, including third-party messaging apps, as well as payment, ride-sharing, and internet calling apps.

- With the release of iOS 11, Apple updated Siri's voices for clearer and more human voices, started supporting follow-up questions and language translation, and additional third-party actions

Figure 3. Siri Architecture
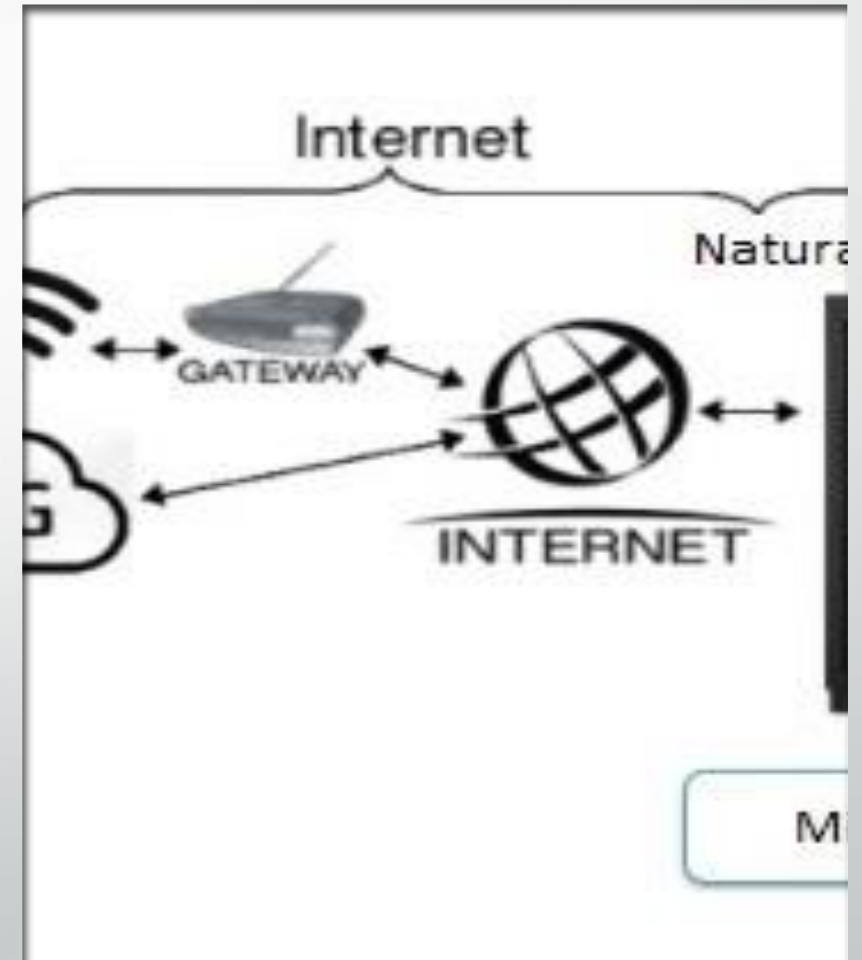
Functional Requirements

According to software engineering, a functional requirement defines a functionality of a component or the entire system. It specifies the services the system must provide, how the system must react to particular inputs, and how the system must behave in particular situations.
- Opening native OS Windows applications: Calculator, Notepad, Paint, etc.
- Opening frequently used applications in Windows OS: Word, Excel, etc.
- Setting a reminder using Sticky Notes
- Search for a file in the file system
- Using a mini voice calculator for simple operations
- Internet browsing
- Accessing web pages: Facebook, Google, UTCN, etc.
- The ability to perform the Login operation on Facebook and Gmail
- Ability to log out on Facebook and Gmail
- Checking the weather
- Use of Google Maps
- Search topics on Wikipedia
- The ability to write and send an email using Gmail.
- Using a transcriber (provide a .wav audio file and identify words)
- Adding or deleting products in a shopping list and listing existing ones in the list.

Non-Functional Requirements

According to software engineering, a non-functional requirement is a requirement by which we can judge the operation of a system. It specifies constraints on the services or functions offered by the system, such as time constraints, constraints on the development process, standards, etc. These requirements are also called quality attributes for a system.

- Performance: It refers to the fact that an application to be performant must have low response and processing times.
- Efficiency (Effciency): It refers to the fact that a system must provide an answer in a short time but using a small number of resources. Thus the voice recognition process is fast using the minimum of resources.

The main characteristics of the systems are:

- Opening certain applications
- Internet browsing
- Performing Login/Logout operations on certain pages
- Sending an email using Gmail
- Adding or deleting products in a shopping list and listing the existing ones in the list.

Conclusions:

Following the analysis of the systems, it can be said that the developed systems are relatively modern, which meet the current open-source standards. It was developed using the latest technologies, open-source, and the structure is well thought out using design patterns. Considering the mentioned, the system brings the following benefits to users:

- - An open-source system available free of charge
- - Minimal use of resources
- - Saving time
- - Increasing productivity
- - User-friendly interface

In order to create a new system, it would be necessary to make certain new contributions: in addition to adding new functionalities, it is also necessary to create a system using only open-source components, thus making the application free and open to the public for development.