ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 6, June 2016

Siri –The Intelligent Personal Assistant

Sheetal Reehal

MCA III Year Student

SIES College Of Management Studies,

University of Mumbai, India.

Abstract –***Siri is an intelligent personal assistant and knowledge navigator which works as an application for Apple's iOS. Talk to Siri as a friend and it can help you get things done – like sending messages, placing calls and reserving a table.***

This paper focuses on what Siri does, how she works, what makes her better than her competitors and what all she still needs to learn to become even stronger.

Keywords – Siri, iOS, Apple Inc.

Introduction –Siri is a virtual assistant with a voice-controlled natural language interface that uses sequential inference and contextual awareness to help perform personal tasks for iOS users. And, like most of the other keytechnological features in Apple's iOS products, SIRI has its roots in federal funding and research. SIRI is an artificial intelligence program consisting of machine learning, natural language processing and a Web search algorithm (Roush 2010). [4]. Spun out of the DARPA-funded CALO project, Siri was initially developed by Dag Kittlaus and his team of SRI International as an iPhone app. Apple bought Siri in 2010. Siri is currently run by Bill Stasior, vice president, Siri, and the technology is integrated into the iPhone, iPod touch, and iPad, and into Apple's new HomeKit home automation framework.

Mostly used by those who cannot use smartphones in a standard way, Apple's Siri began under another name, VoiceOver. VoiceOver is a similar product produced by Siri, Inc. that was designed for patients with Parkinson's disease and the visually impaired. Apple purchased the VoiceOver technology and re-introduced it to the main consumer market as Siri. [5] Siri was re-introduced with a new interface as part of iOS 5. Although it no longer enjoyed the same level of partner integration it had as an independent app, Apple's Siri better integration with built-in services like iMessage, Mail, Calendars, Reminders, Notes, Stocks, Weather, Music, Safari, and a few external services like Google and Wolfram Alpha. It could also tie into Find my Friends.

Siri enables users of Apple iPhone 4S and later and newer iPad and iPod Touch devices to speak natural language voice commands in order to operate the mobile device and its apps. Users can speak commands –and receive audible confirmation from Siri – to send messages, place calls, set reminders, operate iTunes and more.

To use Siri, you don't need to change the way you speak; Siri is a natural language interface. For example, when you want to create a reminder, you simply say something like, "Remind me to walk the dog." Siri responds with a request for specifics such as, "When do you want to be reminded?" You respond with, "Today at 5 PM." Siri dutifully creates the reminder for you. [3] Siri can work across multiple iPhone and iPad apps as needed in order to accomplish its tasks. Siri also supports extended dictation, enabling users to have their words translated into text for use in e-mail and text messages, Facebook status updates, and tweets, note-taking Web searching and similar operations.

Siri also features prominently in the Siri Eyes Free and iOS in the Car technologies from Apple that provide voice command support of a car's audio system or in-vehicle infotainment system.

Technology/Working - Siri is a notable implementation of several technologies: Nuance Communications' voice recognition and text-to-speech (TTS) technology, Siri's artificial intelligence-like (AI) natural language processing engine and backend services (i.e., processing capabilities and access to data and other resources). Perhaps a useful simplification is to suggest that Siri has three layers: voice processing, grammar analysis-context learning engine and services. When a voice file arrives at Apple's data center, the Nuance speech-to-text engine translates the request into text. Nuance has been in the voice technology business since 1994 and, interestingly, was another spin out from the same lab as Siri (SRI International's STAR). It is fair to say that Siri's actual interface is really Nuance technology…but it's the backend magic of Siri that really makes things pop. With Siri, there are no pre-defined ways of requesting Siri do something or answer a question it simply understands what a user wants to do. Importantly, Siri not only understands spoken words, it understands context. Understanding context requires deciphering natural language and then adroitly accessing the resources at Siri's disposal to perform tasks or correctly answer basic or even certain complex questions. Siri also learns. At the user/handset level, Siri has routines that allow Siri to better understand the subtleties of each individual user's accent and voice characteristics. At a macro level, Siri's backend culls through the millions of requests (think: Google Search or Apple's Genius) and finds things to improve upon. For example, when Siri first launched it voiced "Tee X" but within a week it began saying "Texas". What really sets Siri apart, though it's more of design specification than a technology, is Siri's "friendly edginess" and humour—its persona. Siri tries very hard to be witty and very useful. This is very difficult but critical as this "personality" is what has captured the imagination of the market. When merited, Siri delights users with clever, cheeky and laughter-provoking responses. I very much doubt Siri would be getting all of the attention it has if Siri gave accurate but boring responses every time.

How Siri works

Phase 1: voice recognition
It's apparently the easy part, but it's where everything begins, so it can't be trivial. When you give Siri a command, your device collects your analogvoice, converts it into an audio file (it's translated into binary code) and sends it to the Apple servers. The nuances of your voice, the noise around and the local expressions make difficult to get it done right. It's called **Human User Interface** versus the standard Graphical User Interface we are used to. It's important here that, every day, Apple collects millions of queries of people speaking different languages, in different accents, living on different continents. In other words with their actions and mistakes, people are contributing to the largest crowd sourced speech recognition experiment ever tried on earth. Siri app today receives roughly a billion requests per week and Apple states its speech recognition capability has just a 5 percent word error rate. Recently, Apple acquired the speech recognition company Novauris Technologies, a spinoff of Dragon Systems and also hired several speech recognition experts, to get to this point.

Phase 2: send everything to Apple servers in the cloud
Siri does not process your speech input locally on your phone. This is clearly a problem if you're not connected for any reason, but this way Apple gets two major benefits:

- Offload much of the work to powerful computers rather than eating the limited resources of the mobile device

- use the data it collects to continuously improve the service.

The algorithm identifies the keywords and starts taking you down the **flowchart branches** related to those keywords to retrieve your answer. If it fails in this exercise, because a piece of the communication does not work, it goes down the wrong flowchart branch. If it happens just once, the whole query is ruined and ends into the "Would you like to search the web for that?" result. Google Now and Cortana are no different.
You understand this is far from the concept of human conversation. Siri app is still built with a logic of pre-programming all the possible set of questions and rules to answer. This was even more evident when, in October 2015, Apple honoured "Back to the Future" day by updating the Siri app with at least ten humorous responses related to the popular movie Back to the Future. My favourite "be careful who you date today, or you could start disappearing from photos…" is just one answer it picks up randomly from the list.

Phase 3: understand the meaning
The process of understanding what the user is asking for, relies on an area of science called **natural language processing**. People have dozens of ways of asking the same thing. We can express a concept using endless combinations of

words. “I’m in the mood for a pizza”, “Is there an Italian restaurant nearby?”, “I’d love a Margherita today”. Humans can easily understand what I mean, it’s obvious that Margherita is not a person, but an algorithm must be sophisticated to reach the same conclusion. Sometimes it’s just because words have a similar sound or are mispronounced: oyster and ostrich, school and skull, byte and bite, sheep and ship and many others make the task complicated.

To simplify its life, Siri app software, models linguistic concepts. It analyses how the subject keyword is connected to an object and a verb. In other words it looks at the syntactical structure of the text. The decision to go down a branch of the flowchart or another depends upon nouns, adjectives, verbs, as well as the general intonation of the sentences. On top of it, Siri can make sense of questions and follow up commands. This is not exactly what a human would call “a conversation”, but it means it understands the context and it’s the starting point for future developments.

Phase 4: transform the meaning into actionable instructions

We know that Siri is here to help us, not just to understand what we say. In “The story behind Siri”, the founder Adam Cheyer says “I remember the first time we loaded these data sources into Siri, I typed “start over” into the system, and Siri came back saying, “Looking for businesses named ‘Over’ in Start, Louisiana.” “Oh, boy,” I thought.” When the Siri app understands what you want, she has to dialogue with other apps to make it happen. And every app is different and partially has its own “language”. The system must have what is called domain knowledge; it must know the subject area you’re talking about. In a human conversation, this happens every time we talk with experts in a certain field and they use specialized words that we hardly understand. It’s obvious when we speak with a doctor, an architect or a finance person, for example. For the Siri app it’s the same. When it has to give a direction, book a flight or send a text it has to dialogue with other apps… and understand their context. This is crucial as well. If the protocol does not work, Siri can give instructions to other apps to perform actions you didn’t require and expect or can be even potentially dangerous to you. Last but not least, once a request has been processed, Siri must convert the result back into text that can be spoken to the user. While not as hard as processing a user’s command, this task, known as natural language generation, still presents some challenges. Today Siri speaks with the American voice of as “Samantha”, provided by Susan Bennett in July 2005, the same person that voiced Tillie the All-Time Teller. Apple also intends to increase the number of languages that Siri understands. At the outset, Siri understands English (American, British and Australian), Germany and French[2]; and that’s another reason why Siri app is not growing as fast as the original expectation.

Competition –

Google Now:

Launched in 2012, Google Now is an intelligent personal assistant made by Google. It was first included in Android 4.1 which launched on July 9, 2012, and was first supported on the Google Nexus smartphone.

Found within the Google search option, Google Now can be used in numerous ways that are helpful. Yes, it can set reminders or answer basic questions like the weather of the day or the name of the movies that won Oscars last year. But more than that Google Now is a virtual assistant that shows relevant and timely information to you once it learns more about you and how you use the phone.

Google Now also displays different sections called Now cards that pulls information from your Gmail account and throws it on the screen. For example if you have last bought a Red Bag from Amazon, the card shows you your recent buy. Similarly, it also has weather card where you can know about the weather, sport card where you can learn about any match that is on.

Cortana:

Cortana is the name of the interactive personal assistant built into Windows 10**.** You can give her instructions and talk with her by using your voice or by typing.

Cortana, named after her fictional counterpart in the video game series Halo, takes notes, dictates messages and offers up calendar alerts and reminders. But her real standout characteristic and the one Microsoft's betting heavily on, is the ability to strike up casual conversations with users; what Microsoft calls "chitchat."

Facebook arrives:

And now the world’s largest social network has added an IPA voice to its messaging service. Facebook Messenger is testing a new service simply called “M.” David Marcus, head of

Facebook Messenger, explained, “M can actually complete tasks on your behalf. It can purchase items, get gifts delivered to your loved ones, book restaurants, travel arrangements, appointments and way more.” [1]

Conclusion –

In the end, I’d like to conclude my research paper by saying that Siri, without a doubt, is the strongest among its competitors. It can get even better by adopting a few qualities from its rivals. For instance, it could complete the tasks on behalf of the user or it can start learning about the user just like Google Now does and then use this information to complete the user’s tasks.

References –

[1] Michael Castelluccio, “Siri’s Competition”, Technology Workbook, pp.55-56, 2015

[2] David Pogue, “IPhone: The Missing Manual 5th Edition”, O’Reilly Media, pp.104

[3] Brad Miser, “Using IPhone’s Siri Voice Command”, pp. 5

[4] Mariana Mazzucato, “The Entrepreneurial State: Debunking Public vs. Private Sector Myths”

[5] Amy J. Connolly, T. Grandon Gill, “Debates in Information Technology:: A Capstone Class Project”, pp. 112