**Natural Language Processing - Digital Assignment 1**

Name: Ananya Bal

Reg No: 16BCE0063

**ALEXA/ Amazon Lab126**

Amazon Alexa is a virtual assistant developed by Amazon, which is the ultimate embodiment of Natural Language Processing. But you already know that. Everyone has used Alexa.

But you probably don’t know the people responsible for the development of Alexa or just what inspired them. Let’s explore……..

**Inception**

A decade ago, a device that heard you speak and understood you was science fiction, let alone reply back to you. Perhaps, that’s why it seemed apposite in sci-fi movies like Star Trek, where a computer onboard the Starship Enterprise carried out commands **spoken** to it and also spoke back to the crew members of the ship.

But to dreamers like the Indian engineer from Ranchi, Rohan Prasad, who is now the head scientist at Alexa AI, and his colleagues, the computer from Star Trek became an inspiration. And thus, the idea of an intelligent-speech-driven-assistant was conceived at Amazon Lab126 in Sunnyvale, California. Lab126 became home to Alexa and currently tests Alexa devices.

The christening of the technology was a challenge though. Since the name would also double up as the **wake-word** for the device (the word that starts it up – **Alexa,** what’s the weather like?), it had to be short and unique. After testing a bulk of names, the team came up with **Alexa**, it had soft vowels and a distinctive ‘x’. The team also rooted for the name as it was reminiscent of the Library of Alexandria, which is one of the biggest and most comprehensive libraries in the world.

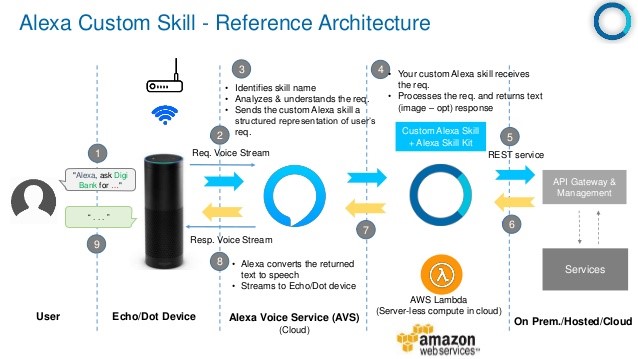
**Design**

Alexa records speech and processes the audio. The audio processing requires the removal of background noises and detection of the direction of audio to better focus on the commands. Seven microphones are used for this purpose.

Post the signal processing, Alexa detects the **wake-word** to activate itself to heed to commands. This is necessary to reduce false positives and false negatives. **Without this, all recorded audio is treated as data to be recorded. Glitches in this wake-word detection have led to Alexa devices recording conversations and compromising the privacy of users. In one case, Alexa had recorded a conversation between a woman and her husband from Portland, Oregon, and had sent it to a random contact of theirs.**

Once the device is active, it transfers the signals to the Amazon Cloud where speech signals are converted to text. **Words are matched with all English words based on phonemes (pitch and frequency of user’s speech are analyzed to give feature values)**. Cloud is the only technology that is able to scale up to the sheer volume and speed required for this matching.

A decoder then determines what the most likely sequence of words is, given the input features and the model. This sequence is split into two parts. The first part is the **prior**, which gives you the most likely sequence based on a large amount of existing text without looking at the features. The other is the **acoustic model**, which is trained with deep learning by looking at pairings of audio and transcripts.



Once the sequence of words is established, **Alexa Voice Service (AVS)** uses NLP to resolve the command into various **Parts-of-speech.** Once the necessary parts are recognized, the contextual meaning is extracted from the command. The **Natural Language** **Understanding that Alexa employs** (statistical machine learning and deep nets) **enables it to do the following with the command: 1. Identify intent 2. Identify utterances (phrases used to refer to a goal) 3. Cover corrections (errors made by the user) and 4. Build exceptions (where Alexa admits she doesn’t know the answer instead of giving a wrong one).**

Entities are extracted, identified, and resolved, and semantic meaning is derived within context, and used for identifying intents. For example, a simple phrase such as: “I need a flight and hotel in Bangalore on December 5” is parsed and given structure:

“**need:flight {intent} / need:hotel {intent} / Bangalore {city} / DEC 5 {date} / sentiment: 0.5723 (neutral)”**

After the goal of the command is found, either the goal is carried out through connected hardware devices or requires Alexa to give back audio output to the user. If the output requires any content like news or weather forecast, the data is searched and the output is designed around it in the opposite process that the input command is decoded, again making use of the complex NLU algorithms that Amazon Voice Service uses.

**Code (AVS API code for Authorisation from a companion app)**

private RequestContext mRequestContext;

private static final String PRODUCT\_ID = "INSERT YOUR PRODUCT ID FROM AMAZON DEVELOPER CONSOLE";

private static final String PRODUCT\_DSN = "INSERT UNIQUE DSN FOR YOUR DEVICE";

private static final String CODE\_CHALLENGE = "INSERT CODE CHALLENGE FROM DEVICE FOR THIS REQUEST";

private static final String CODE\_CHALLENGE\_METHOD = "S256";

@Override

protected void onCreate(Bundle savedInstanceState) {

super.onCreate(savedInstanceState);

mRequestContext = RequestContext.create(this);

mRequestContext.registerListener(new AuthorizeListenerImpl());

*// Find the button with the login\_with\_amazon ID*

*// and set up a click handler*

mLoginButton = (Button) findViewById(R.id.login\_with\_amazon);

mLoginButton.setOnClickListener(new OnClickListener() {

@Override

public void onClick(View v) {

final JSONObject scopeData = new JSONObject();

final JSONObject productInstanceAttributes = new JSONObject();

try {

productInstanceAttributes.put("deviceSerialNumber", PRODUCT\_DSN);

scopeData.put("productInstanceAttributes", productInstanceAttributes);

scopeData.put("productID", PRODUCT\_ID);

AuthorizationManager.authorize(new AuthorizeRequest.Builder(mRequestContext)

.addScopes(ScopeFactory.scopeNamed("alexa:voice\_service:pre\_auth"),

ScopeFactory.scopeNamed("alexa:all", scopeData))

.forGrantType(AuthorizeRequest.GrantType.AUTHORIZATION\_CODE)

.withProofKeyParameters(CODE\_CHALLENGE, CODE\_CHALLENGE\_METHOD)

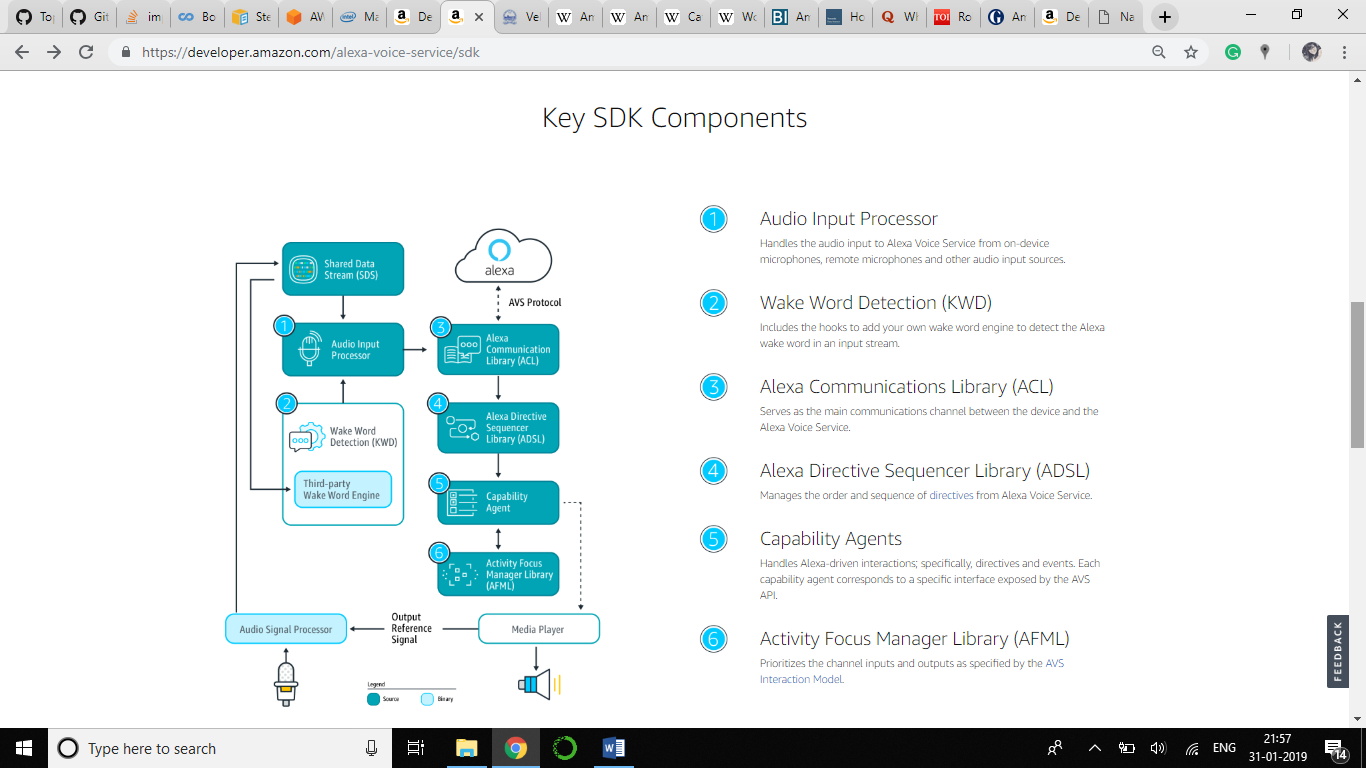
.build());

} catch (JSONException e) { *// handle exception here* }

}

});

}



**Challenges**

The first challenge is the signal filtering, since background noises need to be eliminated and the commanding voice needs to be enhanced. A big problem that had the team worried was the response of the system to the **various accents** that people have, especially in a diverse country such as India. **This influences the signal processing as well as the command resolution using the decoder.** The engineers have revealed that a huge amount of people say “Alexa, news kya hai?” switching to languages they are more comfortable with.

A problem with the **lexicon is the inclusion of names of various famous artists or newsmakers**. Constant additions need to be made to the lexicon to keep it up to date and also make the system inclusive of the various pronunciation of proper nouns like names.

Users can be expecting the weather report with a question like “Alexa, what’s it like outside?” **Alexa needs to resolve the intended context although no keywords like “weather forecast” or “weather” appear in the command.** This requires its NLU algorithms to be robust. This is an ever-growing challenge with slang evolving every day.

And lastly, it goes without saying that privacy is a huge concern for all Alexa users. While any recorded audio after the wake-word is stored by Amazon for improving their AVS model, it is key to not misinterpret the audio as a command which may compromise the user’s privacy.

**My thoughts:**

While Alexa is easily state-of-the-art, it certainly has scope for improvement, and not just in the privacy department. Many of the improvements users demand are hardware or personalization based. However, **I believe that there are certain NLP-centric issues that need to be focussed on**.

1. **Text Summarization** – Not only will this help Alexa to read the news and reviews quicker but also help with its own understanding of intent. Often, human beings speak in long, drawn out sentences. These are characteristic of a human conversation, not a textbook audio command interface with a machine. However, many speak to voice assistants as if they were human. For this reason, having a good summarization capacity with greatly help Alexa become a powerful tool.
2. **More Languages** – The inclusion of more languages, or at least the understanding of basic sentences in more languages, will **help Amazon build a great corpus that many researchers will lust for and will also help improve translation paradigms. Amazon easily has the best infrastructure to handle and learn from massive amounts of multi-lingual linguistics data.**

**Inclusion of more languages will also help eradicate a certain bias that NLP has towards English and the other widely spoken languages in the developed world. Speaking of bias, I believe that having more voices and accents for Alexa will also help.**

**References:**

<https://en.wikipedia.org/wiki/Amazon_Alexa#History>

<https://developer.amazon.com/alexa-skills-kit/nlu>

<https://www.quora.com/What-is-the-technology-behind-Amazon-Echo>

<https://towardsdatascience.com/how-amazon-alexa-works-your-guide-to-natural-language-processing-ai-7506004709d3>

<https://timesofindia.indiatimes.com/nri/us-canada-news/rohit-prasad-the-indian-engineer-who-is-the-brain-behind-alexa/articleshow/65096284.cms>

<https://www.theguardian.com/technology/2018/may/24/amazon-alexa-recorded-conversation>