MINI PROJECT

Submitted by

DUSHYANT RAO [RA2011028010106] AVIPSHA PANIGRAHI [RA2011028010101] SHUBHRA KUMARI [RA2011028010093]

Under the Guidance of

Mrs. D. Saveetha

Assistant Professor DEPARTMENT OF NETWORKING AND COMMUNICATION

In partial satisfaction of the requirements for the degree of

BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE ENGINEERING with specialisation in CLOUD COMPUTING

SRM INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

SCHOOL OF COMPUTING COLLEGE OF ENGINEERING AND TECHNOLOGY SRM INSTITUTE OF SCIENCE AND TECHNOLOGY KATTANKULATHUR – 603203 NOVEMBER 2022

SRM INSTITUTION OF SCIENCE AND TECHNOLOGY KATTANKULATHUR-603203

BONAFIDE CERTIFICATE

*This is to certify that the project titled **"Universal Translator Using AWS "** is the bonafide work of **DUSHYANT RAO [RA2011028010106], AVIPSHA PANIGRAHI [RA2011028010101], SHUBHRA KUMARI [RA2011028010093]** of **Computer Science and Engineering** B.Tech degree course who carried out the project under my supervision.*

Signature of the Course Faculty
Mrs. D. Saveetha
Assistant Professor
Department of NWC, SRM Institute of Science and Technology

ABSTRACT

Universal translator is a project that aims to provide translation services for cloud based applications. It uses a combination of machine translation to provide accurate and reliable translations. The project is still in its early stages, but it has the potential to become a valuable tool for developers and users of cloud-based applications.

The purpose of this project is to create a cloud-based solution using Amazon Web Services (AWS). The project will use AWS resources to create a virtual private cloud (VPC).

The web application will be accessible to users via a web browser. The project will also use Amazon Simple Storage Service (S3) to store data for the web application.

In recent years, the use of cloud-based machine translation has increased exponentially, with Amazon Web Services (AWS) being one of the most popular platforms. Amazon Translate is a neural machine translation service that uses deep learning algorithms to provide high-quality translations.

In this project, we will investigate the use of Amazon Translate for the translation of Any language texts into any other languages. We will also evaluate the quality of the translations produced by Amazon Translate and compare it to that of other machine translation services.

How we are going to achieve these things-

When a new document is uploaded to a specific S3 bucket i.e input S3, we configure the setting on this S3 bucket to send a notification to AWS Lambda. This notification AWS Lambda to run a code to performs the following sequence of events –

1)Read the document upload to the Input S3 bucket. Extract short segments from the document

3)Pass these segments through the Real-Time Translation API.

4)Use the Real-Time Translation API's output to rebuild the translated output document

5)Save the output in the specified output location. i.e Output S3 bucket

Therefore at the end we would like to conclude that-

AWS (AWS) is a cloud computing platform that offers users a wide range of services including storage, networking, analysis, and more. AWS is one of the most popular cloud platforms on the market, and is used by major companies such as Netflix, Airbnb, and Facebook.

TABLE OF CONTENTS

Literature Survey

This paper describes a machine translation system that is capable of translating a grammatically correct Sinhala sentence into its corresponding English sentence.

This is the first Sinhala to English machine translation system, which comes with features such as an inbuilt keyboard, an inbuilt dictionary, an integrated word processor based on Unicode fonts, a grammar tool, a Sinhalese grammar checker, an add word tool, and a debugging tool.

With the expansion of the world, English has become an important language that people should learn, as the majority of the worldwide population understand and carry out their day-to-day work in English.

In addressing this need, we thought of taking up the challenge of building a Sinhala to English language translator. To build this system, we used the Transfer-based machine translation approach, which is a Rule-based approach. At present, the system has achieved a success rate of 75% with a corpus of 150 sentences.

Comparative Analysis of Language Translation and Detection System Using Machine Learning

August 2021
International Journal of Scientific Research in Computer Science Engineering and Information Technology **Abstract:**

Words are the meaty component which can be expressed through speech, writing or signals. It is important that the actual message or meaning of the words sent must convey the same meaning to the one received. The evolution from manual language translator to the digital machine translation has helped us a lot in finding the exact meaning such that each word must give at least close to exact actual meaning.

To make machine translation more human-friendly, natural language processing (NLP) with machine learning (ML) can make the best combination.

The main challenges in machine translated sentences can involve ambiguities, lexical divergence, syntactic, lexical mismatches, semantic

issues, etc. which can be seen in grammar, spellings, punctuations, spaces, etc.

After analysis on different algorithms, we have implemented two different machine translations using two different Long Short-Term Memory (LSTM) approaches and performed the comparative study of the quality of the translated text based on their respective accuracy.

We have used two different training approaches of encoding decoding techniques using the same datasets, which translate the source English text to the target Hindi text.
To detect whether the text entered is English or Hindi language, we have used a Sequential LSTM training model for which the analysis has been performed based on its accuracy. As the result, the first LSTM trained model is 84% accurate and the second LSTM trained model is 71% accurate in its translation from English to Hindi text, while the detection LSTM trained model is 78% accurate in detecting English text and 81% accurate in detecting Hindi text. This study has helped us to analyze the appropriate machine translation based on its accuracy.

Keywords: Accuracy, Decoding, Machine Learning (ML), Detection System, Encoding, Long Short-Term Memory (LSTM), Machine Translation, Natural Language Processing (NLP), Sequential

INNOVATIVE IDEA

Innovative idea in our project is that we are using different services in order to translate a particular text document. The user will upload the **text document** into the input S3 bucket and it can download the translated text document from the output S3 bucket. So the innovative idea used here is we are directly converting the whole text document which is in another language into a user specific language.

The whole **process is done in one go** so there is no need to copy paste text. That will save a lot of time. **Punctuation symbols , lines , paragraphs are not being changed while in the process of translating so the user will get an absolute copy of the text file in the desired language in a perfect format.**

Here format is the most important aspect of our innovation. In a traditional translator the user has to copy and paste text line by line in order to translate, but here all the paragraphs, all the full stops, commas and the whole format of that particular text document is not being disturbed.

We are trying to make a universal Translator using Amazon web services (AWS).

WHICH CAN TRANSLATE ANY LANGUAGE TO SPECIFIC LANGUAGE.

1. Try to translate as many different types of language as possible.

2. Try to make this project by using FREE tier services provided by AWS. (i.e cost efficient)
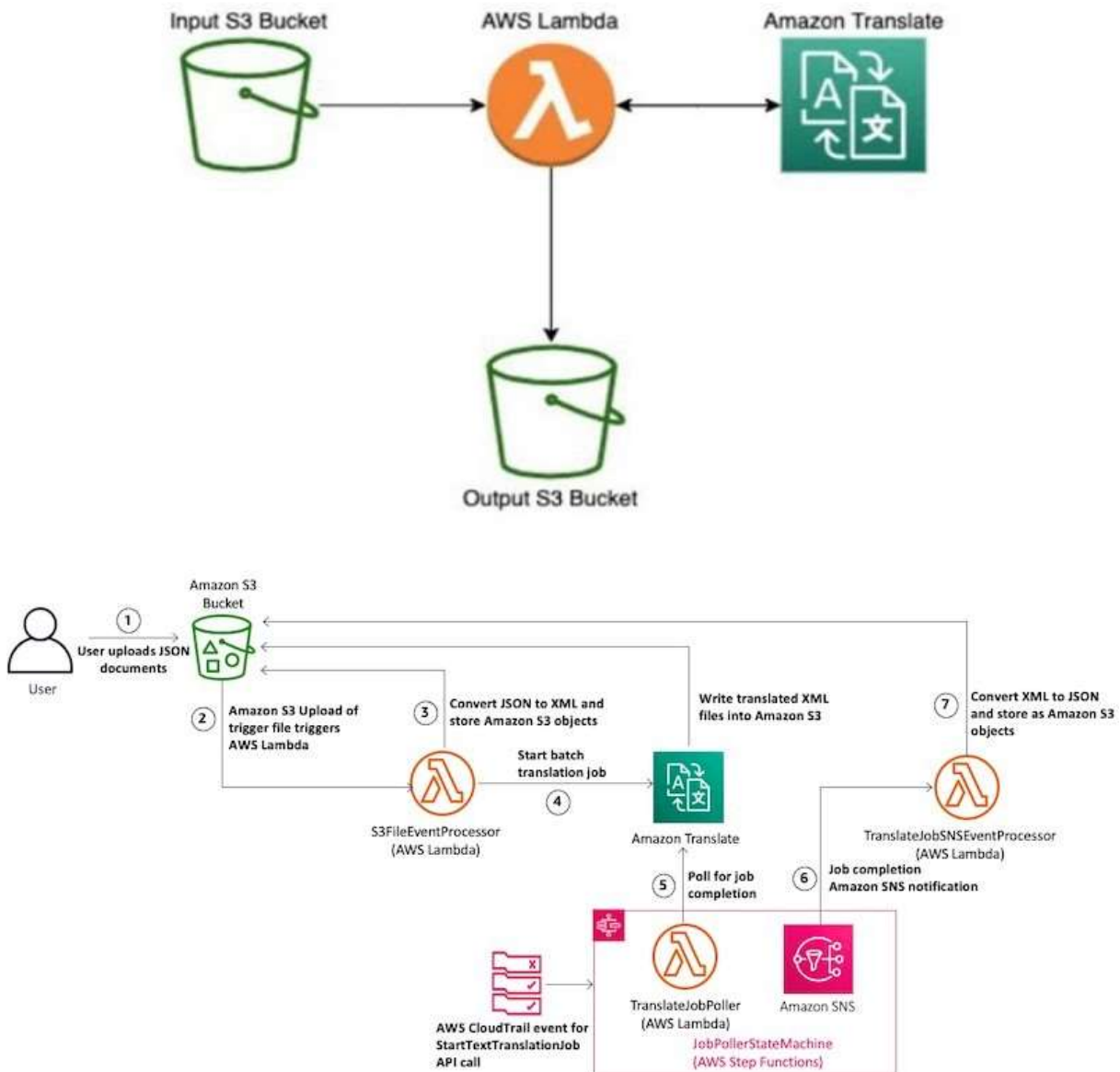
SYSTEM DESIGN

Architecture Diagram

We are using **Event-Driven Architecture** in our project
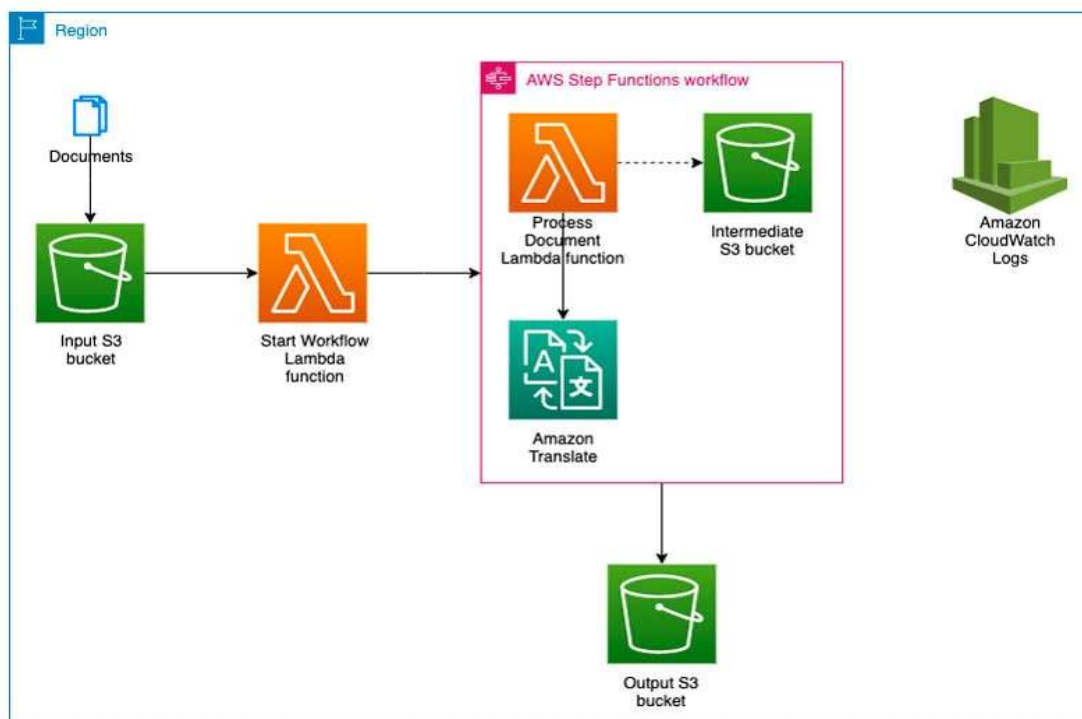
MODULE

Here in this module the basic working of of our project is being shown, input S3 bucket is used to Trigger the Lambda function then the code in the Lambda will send the user uploaded text document line by line to Amazon translate and the translate will convert it into the suitable language and give it back to lambda after that the user can download that translated document from output bucket.

# IMPLEMENTATION

Technologies Used

AWS S3 bucket

An S3 bucket is a logical unit of storage in Amazon S3. All objects stored in Amazon S3 must belong to a bucket. Buckets serve as a top-level container in Amazon S3, which stores objects. Amazon S3 has a flat structure with no hierarchy like you would see in a typical file system. However, for the sake of organization, the Amazon S3 console supports the folder concept as a means of grouping objects. An Amazon S3 bucket name is globally unique, and the namespace is shared by all AWS accounts.

This means that once a bucket name has been created, the name cannot be used by another AWS account in any region, unless that account is the owner of the bucket. A typical Amazon S3 bucket name would look like this:

my-bucket-name

> However, bucket names cannot contain uppercase characters, and they must be between 3 and 63 characters long.
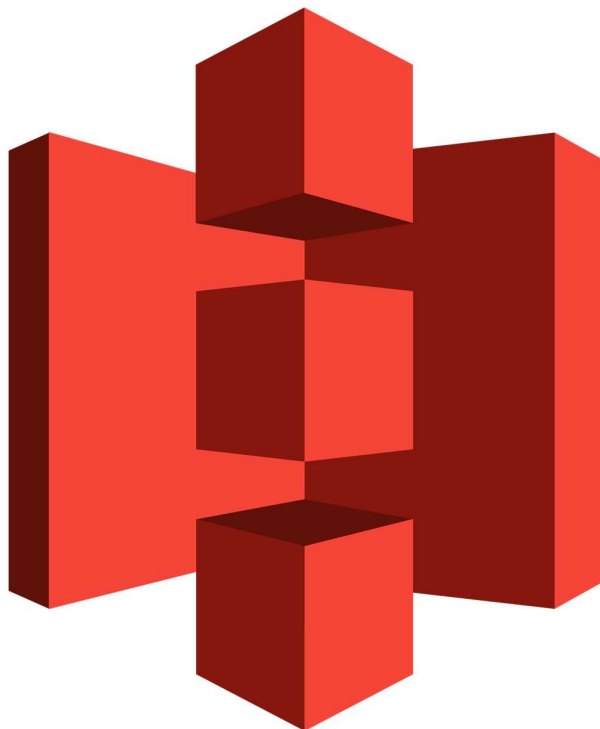
AWS Lambda

AWS Lambda is a serverless computing platform that allows you to run code without provisioning or managing servers.

Lambda handles all the administration of the underlying infrastructure, including server and operating system maintenance, capacity provisioning, and code monitoring and logging.

With Lambda, you can run code for virtually any type of application or backend service - all with zero administration.

> Just upload your code and Lambda takes care of everything required to run and scale your code with high availability. You can set up your code to automatically trigger from other AWS services or call it directly from any web or mobile app.

Amazon Translate

AWS Translate is a neural machine translation service that delivers fast, high-quality translation of documents and websites. Neural machine

translation is a form of artificial intelligence that uses deep learning algorithms to translate text from one language to another.

Every specific language has its own code in AWS translate which only Amazon translate understands.Here are some language codes-

Language	Language code
Afrikaans	af
Albanian	sq
Amharic	am
Arabic	ar
Armenian	hy
Azerbaijani	az
Bengali	bn
Bosnian	bs
Bulgarian	bg
Catalan	ca
Chinese (Simplified)	zh
Chinese (Traditional)	zh-TW
Croatian	hr
Czech	cs
Danish	da
Dari	fa-AF
Dutch	nl
English	en
Estonian	et
Farsi (Persian)	fa
Filipino, Tagalog	tl
Finnish	fi
French	fr
French (Canada)	fr-CA
Georgian	ka
German	de

Amazon Translate

CloudWatch

CloudWatch is a monitoring service for AWS resources and the applications you run on AWS. CloudWatch monitors AWS resources such as Amazon EC2 instances, Amazon DynamoDB tables, and Amazon RDS DB instances, as well as custom metrics generated by your applications and services, and any log files your applications generate. You can use CloudWatch to collect and track metrics, collect and monitor log files, set alarms, and automatically react to changes in your AWS resources.

CloudWatch Logs CloudWatch Logs is a feature of CloudWatch that lets you monitor, store, and access log files from Amazon EC2 instances, AWS CloudTrail, and other sources. With CloudWatch Logs, you can monitor your logs, in near real-time, for specific phrases, values, or patterns (metrics). You can also monitor and watch log files stored in Amazon S3.
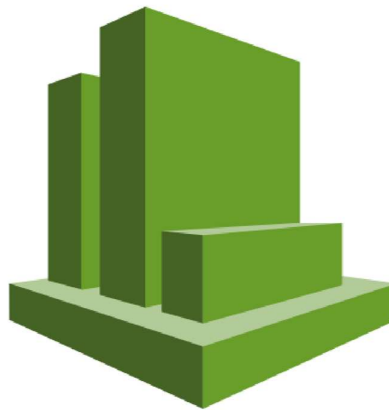
Identity & Access management (IAM)

AWS IAM is a web service that helps you securely control access to your AWS resources. IAM is a feature of your AWS account that is separate from other services such as Amazon EC2. IAM can also be used with other AWS services such as Amazon S3 and Amazon Glacier. IAM enables you to create and manage users and groups and use permissions to allow and deny their access to AWS resources. IAM is a free service that is available to all AWS customers.

AWS Identity and Access Management (IAM) is a web service that helps you securely control access to AWS resources. IAM is a feature of your AWS account offered at no additional charge. You use IAM to control who is authenticated (signed in) and authorized (has permissions) to use resources in your AWS account. For example, you can use IAM to control who can create, update, or delete Amazon EC2 instances or S3 buckets, or who can have readonly access to Amazon DynamoDB tables.
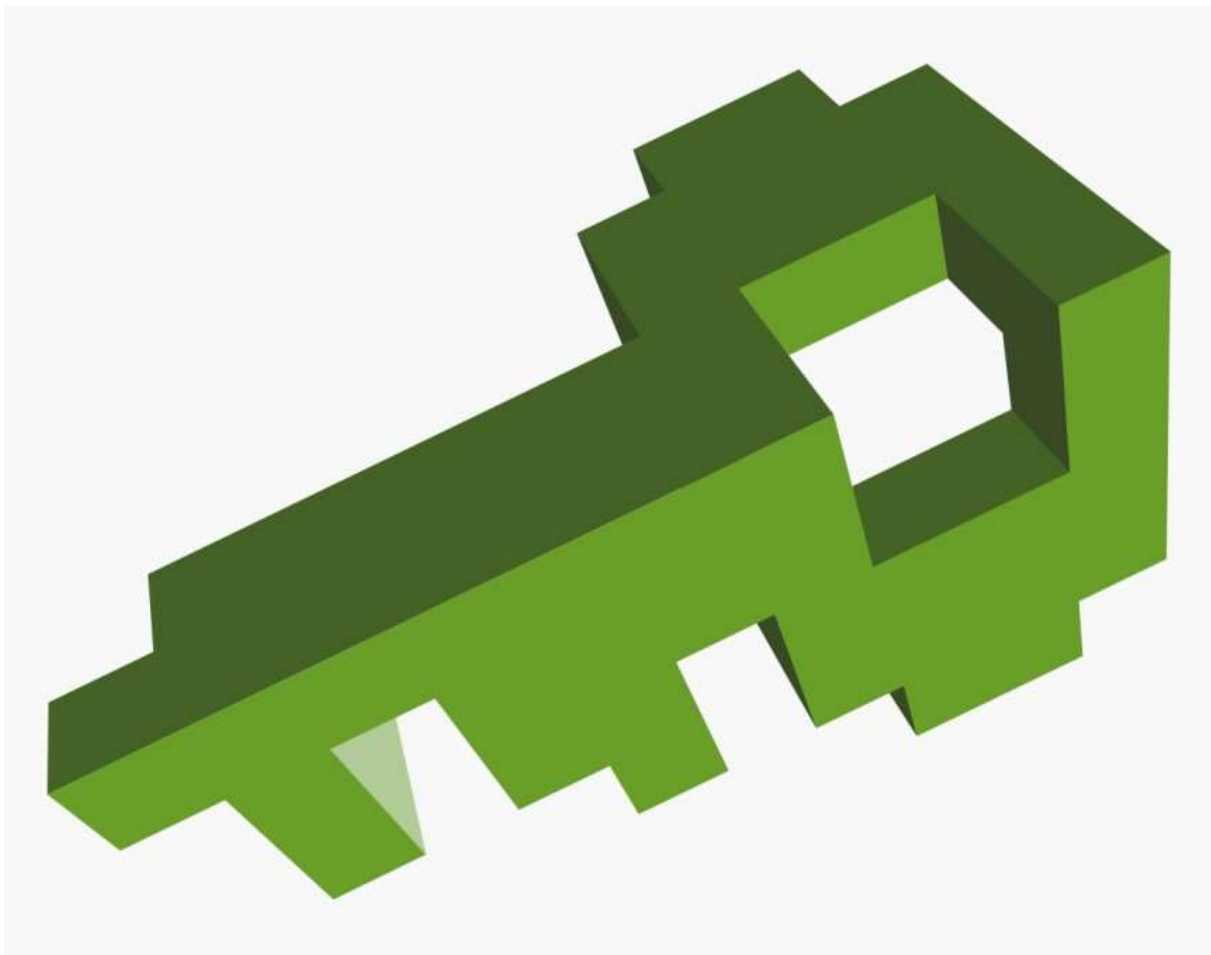
CODE

This piece of code is used to automate the above procedure above.

```python
import json
import boto3

s3_client = boto3.client(service_name='s3')
translate = boto3.client('translate')


def translate_text(text, lang_code):
    result = translate.translate_text(
        Text=text,
        SourceLanguageCode='auto',
        TargetLanguageCode=lang_code
    )
    return result['TranslatedText']


def lambda_handler(event, context):
    file_name = event['Records'][0]['s3']['object']['key']
    bucketName=event['Records'][0]['s3']['bucket']['name']
    outfile="s3://outputbuk/{}".format(file_name)
    print("Event details : ",event)
    print("Input File Name : ",file_name)
    print("Input Bucket Name : ",bucketName)
    print("Output File Name : ",outfile)
    # get S3 object
    result = s3_client.get_object(Bucket=bucketName, Key=file_name)
    #Read a text file line by line using splitlines object
    final_document_array = ""
    for line in result["Body"].read().splitlines():
        each_line = line.decode('utf-8')
        print("Input Line : ",each_line)
        if(each_line!=''):
            translated=translate_text(each_line, 'Hi')
            print("After translation : ",translated)
            final_document_array+=translated
            final_document_array+='\n\n'
    s3_client.put_object(Body=final_document_array, Bucket='outputbuk', Key=file_name)
    print("Done")
```

CONCLUSION AND FUTURE ENHANCEMENT

So, at the conclusion of our project the document uploaded by the user is being successfully translated into user specific format and all the special

symbols format are not being disturbed. The exact format which is being uploaded by the user is being displayed over the output.

For the future and its mint we have ideas like we will give the user privilege to change their desired specific language. Which the user wants to translate the document through front end. Giving them the privilege to change the language according to their convenience. We will provide them with a list of language codes which are recognized by Amazon translation.

Next we will try to make our project the universal translator that accepts document types other than text format, that is, in future our translator will accept file formats like PDF, docs files ,etc.

REFERENCES

https://docs.aws.amazon.com/IAM/latest/UserGuide/id_users_create.html

https://aws.amazon.com/blogs/machine-learning/create-a-serverlesspipeline-to-translate-large-documents-wit h-amazon-translate/ https://docs.aws.amazon.com/index.html