Project Proposal: Contract Comprehension Application

Abstract:

The goal of this project is to create a serverless environment in AWS which allows for the ingestion of contract documents in PDF format, performs a text recognition process, and uses Amazon Comprehend to extract entities and key phrases. The resultant information would be ingested into an AWS elasticsearch (now opensearch) database and visualized through a Kibana (no opensearch dashboard).

This architecture is not exotic and there are examples of how this has been done available. My goals are to learn this content for my own benefit, performing the work myself, but with some coaching along the way. I’m including an architecture diagram and some screenshots of where I am today.

AWS Sample that I’d like to follow (as a place to start): <https://github.com/aws-samples/amazon-comprehend-doc-search>

Diagram

Description automatically generated

Screenshots of various modules that I’ve created so far:

**OpenSearch service:**

Graphical user interface, application

Description automatically generated

Table

Description automatically generated with low confidence

Access Policy:

{

"Version": "2012-10-17",

"Statement": [

{

"Effect": "Allow",

"Principal": {

"AWS": "arn:aws:iam::926494140610:role/text-analysis-with-amazon-ComprehendOpenSearchServ-SX93XH4QVC3M"

},

"Action": "es:\*",

"Resource": "arn:aws:es:us-east-1:926494140610:domain/contract-comprehension/\*"

},

{

"Sid": "",

"Effect": "Allow",

"Principal": {

"AWS": "\*"

},

"Action": "es:\*",

"Resource": "arn:aws:es:us-east-1:926494140610:domain/contract-comprehension/\*",

"Condition": {

"IpAddress": {

"aws:SourceIp": "75.69.20.134"

}

}

}

]

}

**S3 Buckets:** 001 is ingestion bucket. Have been testing outputs with 002 bucket.

Graphical user interface, text, application, email

Description automatically generated

**LAMBDA FUNCTION: Code in Github**

Graphical user interface, text, application, email

Description automatically generated

**Lambda role in IAM:**

Text, table

Description automatically generated with medium confidence