Advanced Text Analysis



Mark Nunnikhoven AWS COMMUNITY HERO @marknca markn.ca

Overview

Examine the concept of a "job"

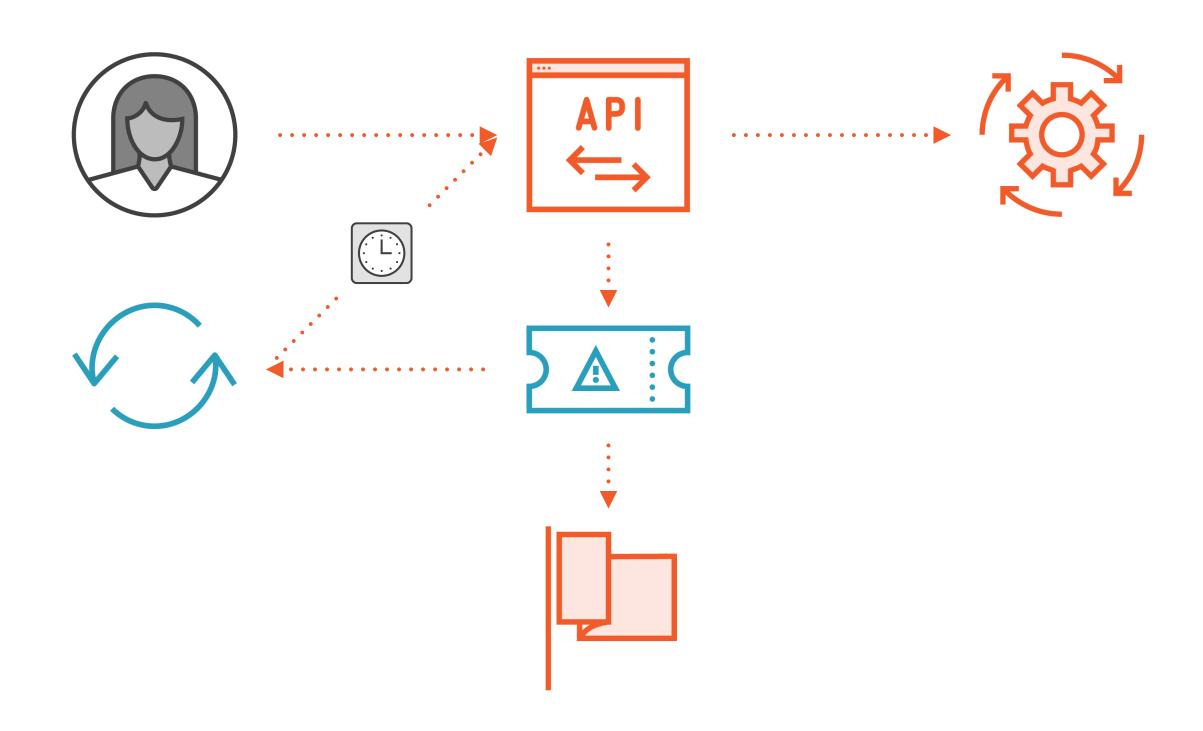
Detecting entities (people, places, etc.)

Augmenting the service with custom entities

Determining key phrases within the text

Asynchronous API Calls with Jobs

The Asynchronous Concept



Asynchronous Calls



- Post text to Amazon S3 bucket
- 2. Receive [S3 URI]



- 3. Call Start___Job using [S3 URI]
- 4. Receive [Jobld]

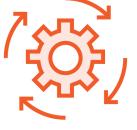




- 5. Call Describe____Job using [JobId]6. Receive [JobStatus]



Repeat #5/6 until [JobStatus] == "COMPLETED"





- 7. Get text from Amazon S3 bucket using [OutputDataConfig]
- 8. Receive the analysis result



Request Format

```
"DataAccessRoleArn": "string",
"InputDataConfig": {
  "S3Uri": "string"
"JobName": "string",
"OutputDataConfig": {
  "S3Uri": "string"
```

▼ The IAM Role used to grant permissions to the service

■ Input document in an S3 bucket

■ Name of the job (optional)

■ S3 path to store the output



Amazon Comprehend needs to be able to read the specified S3 input key (a/k/a Uri)

The service also needs to be able to write the specified S3 output key (a/k/a Uri)

A reasonable compromise is to use two dedicated buckets for the service. One for input and the other for output

Demo

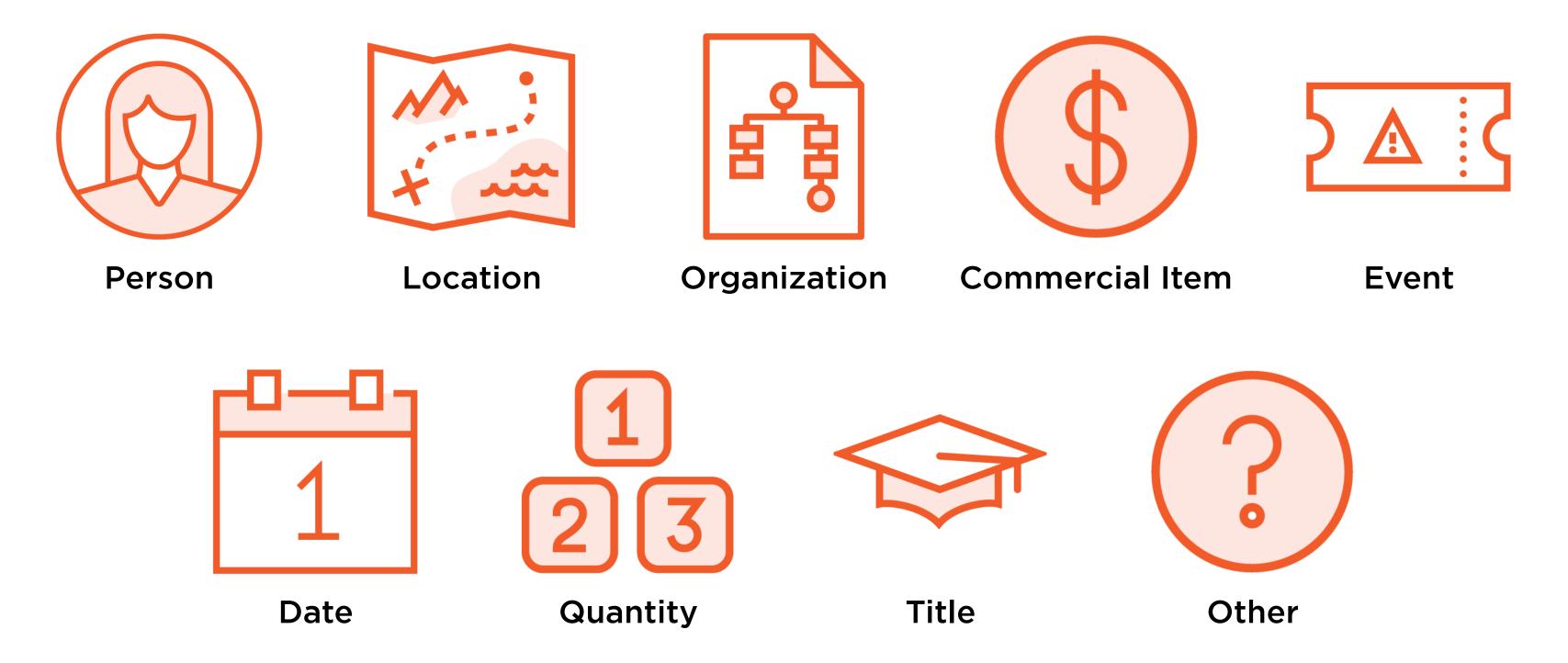
Use the AWS IAM service to create a Role

Assign permissions to that Role

Record the Role ARN for future use

Detecting Entities

Entities



Entity Example

Jim works for the United Nations in New York, New York.



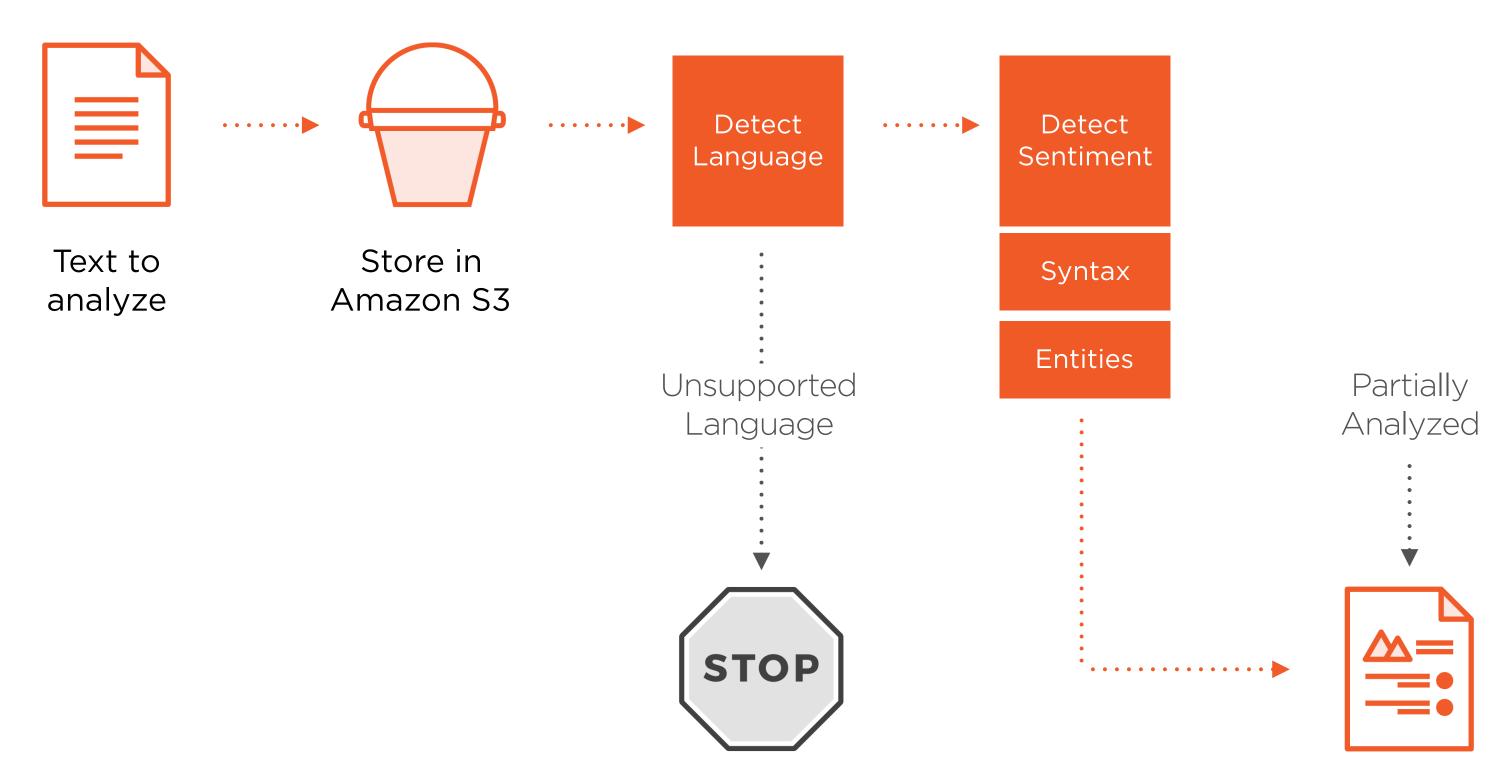




He started there on January 1st, 2019



Document Analysis



DetectEntities

A synchronous API call that locates and categorizes entities within the specified text

BatchDetectEntities

A synchronous API call that locates and categorizes entities for up to 25 samples of text in the same dominant language

StartEntitiesDetectionJob

An asynchronous API call that locates and categorizes entities within a specified document stored in Amazon S3

Asynchronous Calls



- Post text to Amazon S3 bucket
- 2. Receive [S3 URI]



- 3. Call Start___Job using [S3 URI]
- 4. Receive [Jobld]

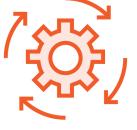




- 5. Call Describe____Job using [JobId]6. Receive [JobStatus]



Repeat #5/6 until [JobStatus] == "COMPLETED"





- 7. Get text from Amazon S3 bucket using [OutputDataConfig]
- 8. Receive the analysis result



Response Format

```
"Entities": [
      "BeginOffset": number,
      "EndOffset": number,
      "Score": number,
      "Text": "string",
      "Type": "string",
   },
```

- **◄** Starting character in the text
- **◆** Confidence score of the classification
- Text classified
- Entity type (e.g., PERSON, DATE, EVENT, etc.)
- **◄** More entities...

Demo

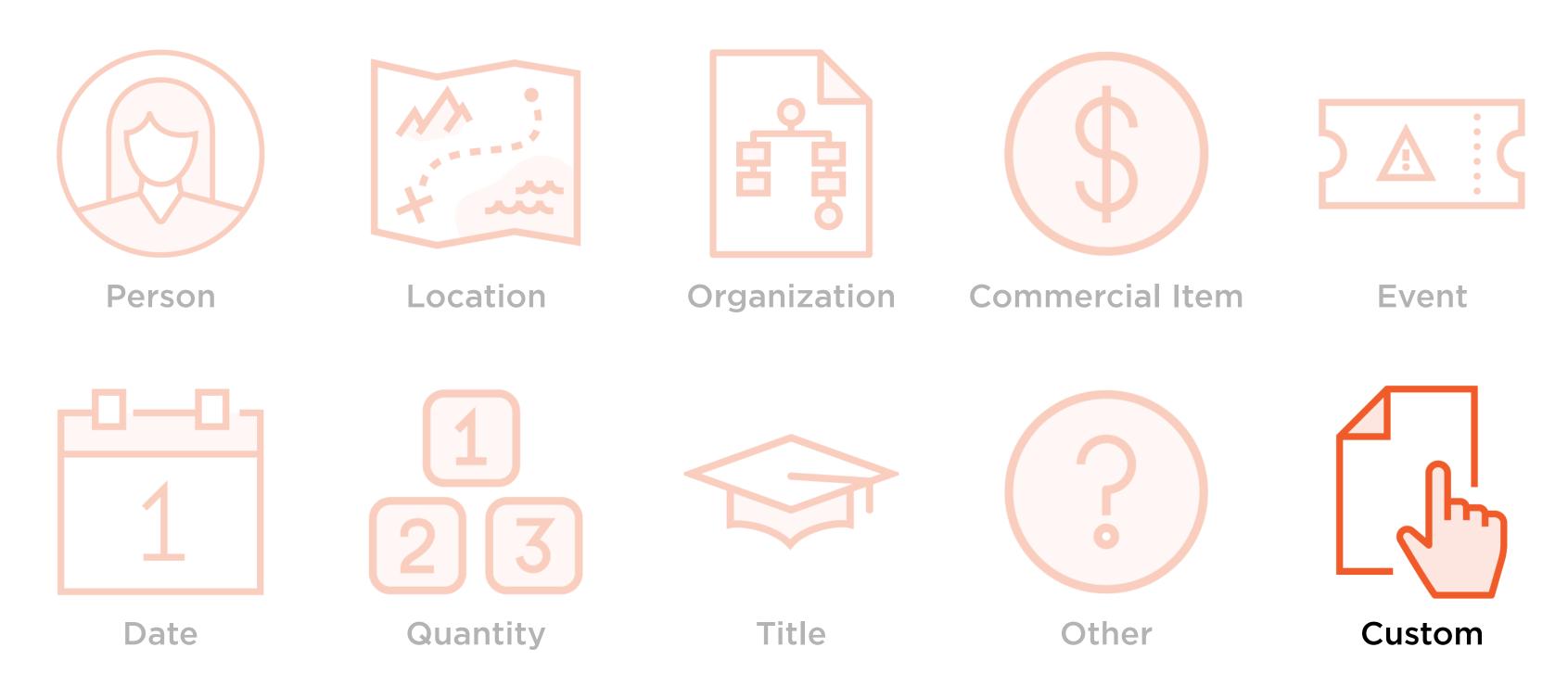
Detect entities using an asynchronous call via code

Poll the API for the job results

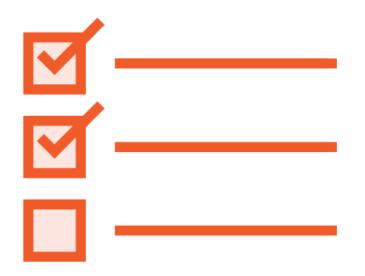
Review the results when the job is complete

Custom Entities

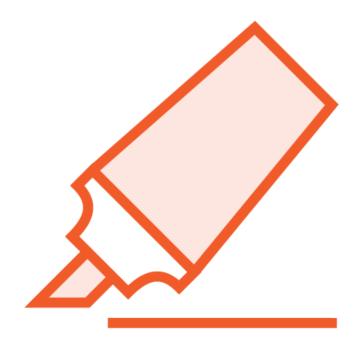
Entities



Custom Entity Recognizers







Annotations

Custom Entity List

Text Type

PluralSight

ORGANIZATION

CreateEntityRecognizer

API

ListEntitiesDetectionJobs

API

EntitiesDetectionJobProperties

DATA_STRUCTURE

Custom Annotations

File	Line	Begin Offset	End Offset	Type
doc1	O	O	10	ORGANIZATION
doc1	1	24	56	API
doc2	1	12	43	API
doc2	2	2	41	DATA_STRUCTURE

Demo

Create an Entity Recognizer using an Entity List

Use the new recognizer

Review the results

Detecting Key Phrases

Key phrase detection locates noun phrases in a text to identify the subject(s)

Key Phrase Example

Detects the key noun phrases found in the text.

0.983156

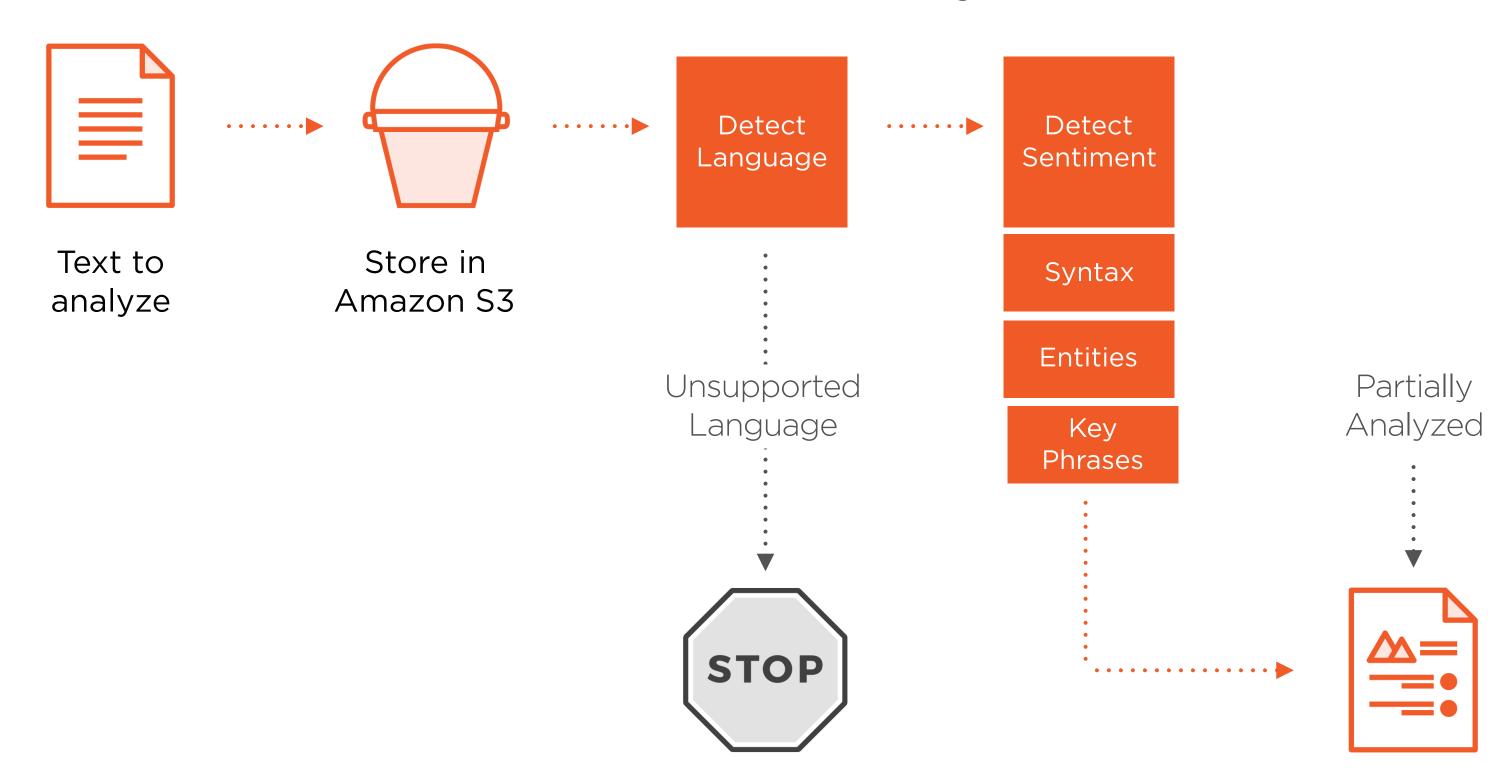
0.999571

This classifier offers all three API methods.

0.986743

0.986447

Document Analysis



DetectKeyPhrases

A synchronous API call that locates key phrases within the specified text

BatchDetectKeyPhrases

A synchronous API call that locates key phrases for up to 25 samples of text in the same dominant language

StartKeyPhrasesDetectionJob

An asynchronous API call that locates key phrases within a specified document stored in Amazon S3

Response Format

```
"KeyPhrases": [
      "BeginOffset": number,
      "EndOffset": number,
      "Score": number,
      "Text": "string",},
```

- **■** Starting character in the text
- **◄** Confidence score of the classification
- Text classified
- **◄** More key phrases...

Demo

Detect key phrases using an asynchronous call via code

Poll the API for the job results

Review the results when the job is complete

Review



Classifiers also provide an asynchronous method of making a call using a "Job". This method allows for large documents to be analyzed



Entity detection highlights the people, places, organizations, events, and other key items of interest in a text



Custom entities allow detection of specific terms that match your use case. This feature is very handy for niche analysis



Key phrase detection highlights the noun phrases within a given text, helping identify objects of importance