



---

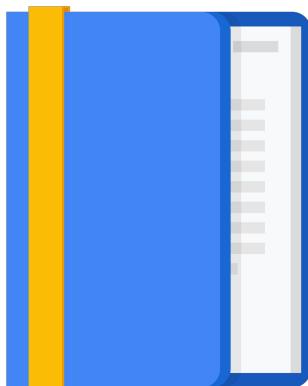
Prebuilt ML model APIs  
for Unstructured Data

# Agenda

---

Unstructured Data is Hard

ML APIs for Enriching Data



How do you extract anything useful out of this?



or



<https://pixabay.com/photos/newspaper-old-1925-daily-newspaper-2608394/>

## Examples of real-world use cases of unstructured data

Custom image  
model to price  
cars

Build off NLP  
API to route  
customer emails

Use Vision API  
as-is to find text  
in memes

Use Dialogflow  
to create a new  
shopping  
experience



Uniqlo (uni-clo) designed a shopping chatbot using A Dialogflow UI. Dialogflow is a google-owned company which specializes in building ML-based interfaces like intelligent chatbots.

Let's look at some other use cases for ML in business

[Resource]

Google NeXT Customer Success Stories  
[https://youtu.be/BwWg\\_\\_HVfsM?t=4m41s](https://youtu.be/BwWg__HVfsM?t=4m41s)

Clouds or snow-capped mountains?



 Google Cloud

You might be Airbus, and use machine learning to differentiate between clouds and snow cover. If you're stumped like I am, the clouds are in the upper-right part of the right image highlighted in red.

<https://cloud.google.com/blog/products/gcp/google-cloud-machine-learning-now-open-to-all-with-new-professional-services-and-education-programs>

## Empty or Full?



 Google Cloud

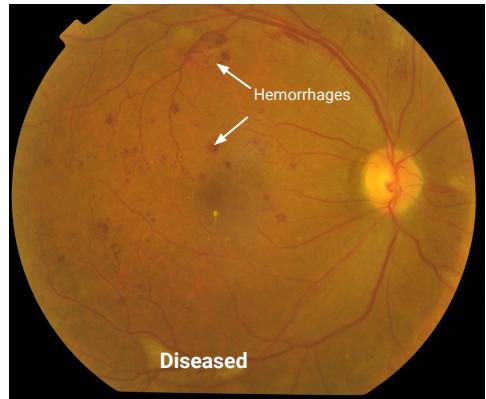
You might be an economic forecast firm looking to track the global fleet of container ships via satellite imagery -- knowing the amount of cargo being carried might help improve your economic forecast days or months ahead of the official numbers.

<https://pixabay.com/en/container-ship-container-ship-port-2786842/>

## Diagnosing Diabetic Retinopathy



Healthy



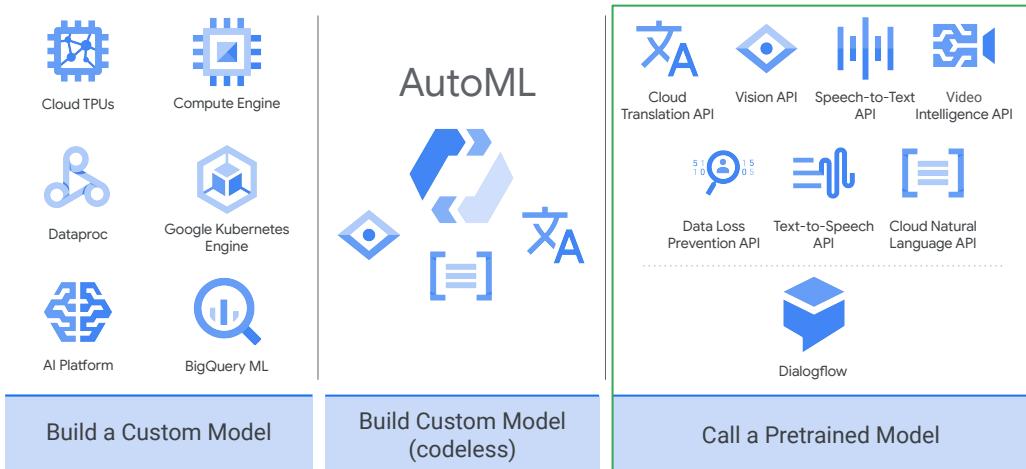
Diseased



Medical images are ripe for innovation. For example, You could diagnose [medical conditions](#) like Diabetic Retinopathy earlier when it's easier to treat and prevent blindness.

<https://ai.googleblog.com/2016/11/deep-learning-for-detection-of-diabetic.html>

## For common ML tasks, consider Pretrained APIs

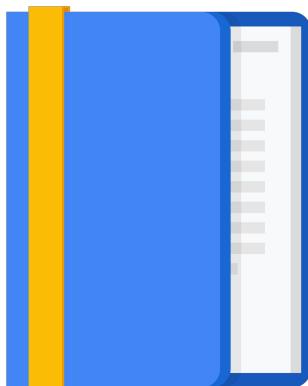


# Agenda

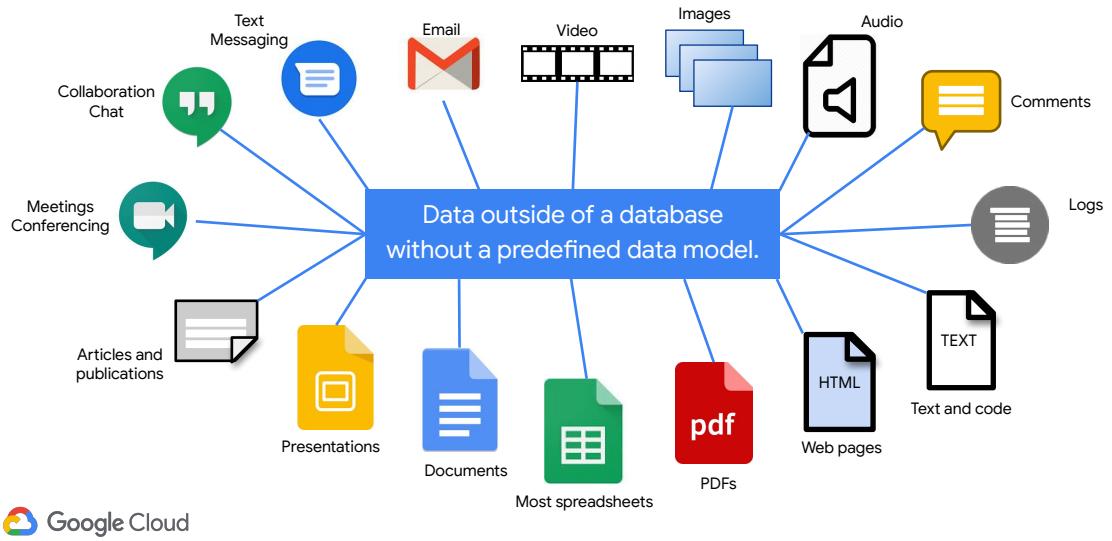
---

Unstructured Data is Hard

ML APIs for Enriching Data

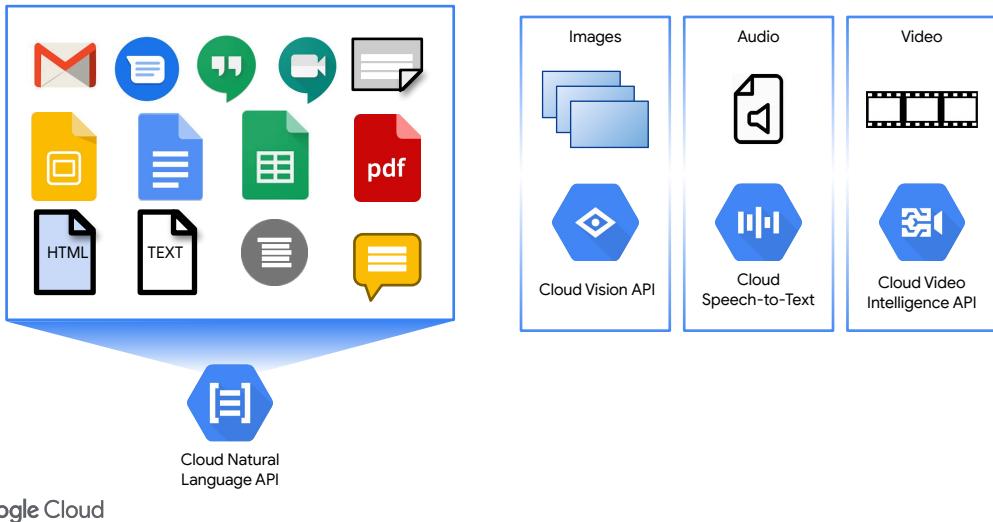


## Most business data is unstructured data, and mainly text



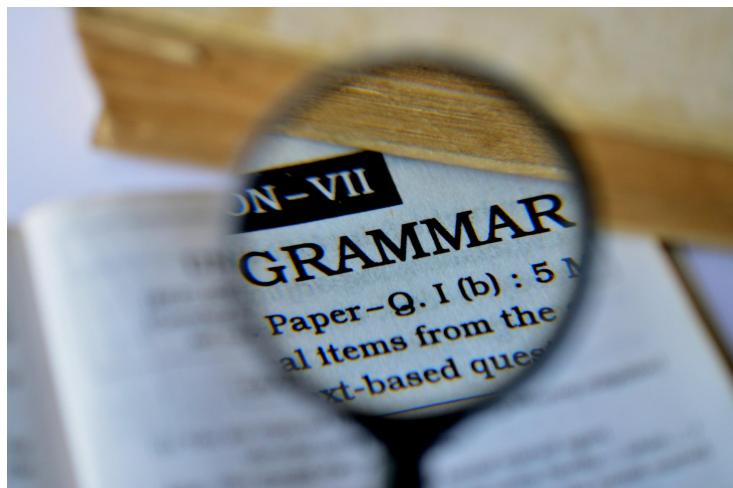
What is structured data? Data in a database, in JSON files, and in some spreadsheets.

You process unstructured data by labeling it with AI



We are going to concentrate on the Cloud Natural Language API.  
It is accessible to you in many ways, but if you already have data that you are  
migrating from Hadoop to Cloud Dataproc, you can make this available for analysis  
using the Cloud Natural Language API.

Syntactic Analysis provides linguistic information



 Google Cloud

Syntactic Analysis breaks up the text into a series of tokens (generally, words and sentences) and provides information about the token's internal structure (morphology) and its role in the sentence (syntax). It can label a token as noun or verb, singular or plural, first person or second person, masculine, feminine, or neutral, and provides grammatical information such as case, tense, mood and voice.

At the time of this writing it supports 10 languages. See online documentation for current language support.

<https://pixabay.com/photos/grammar-magnifier-magnifying-glass-389907/>

## Entity Analysis recognizes many useful items



Locations and landmarks



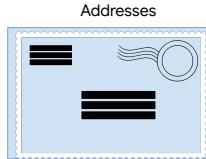
People



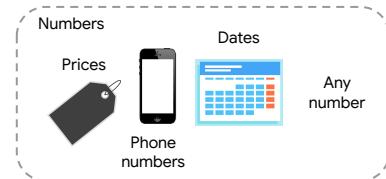
Artwork and works of art



Consumer Products



Addresses



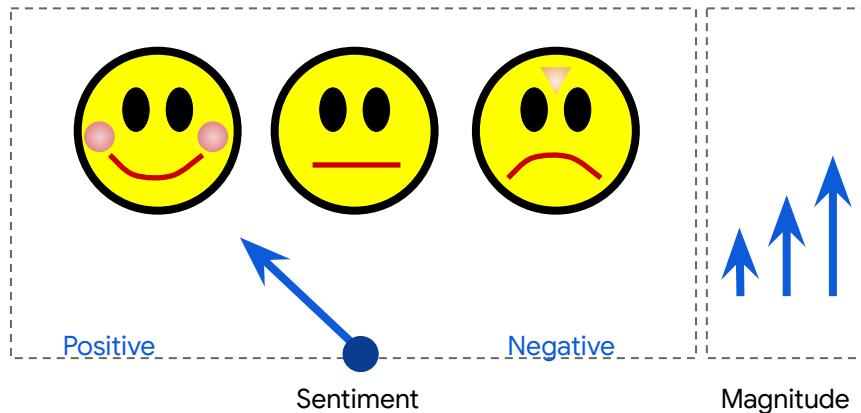
Google Cloud

proper nouns such as public figures, landmarks

Person, Location, Organization, Event, Artwork, Consumer Product, Phone Number, Address, Date, Any number, Price.

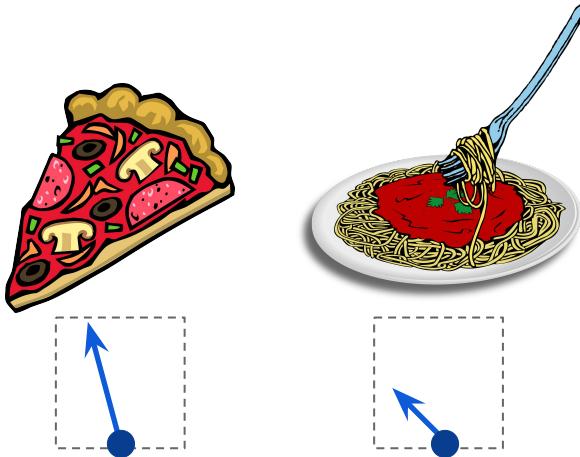
- <https://pixabay.com/illustrations/males-3d-model-isolated-3d-model-2373729/>
- <https://pixabay.com/photos/art-creative-hand-painting-paper-1868562/>
- <https://pixabay.com/photos/eiffel-tower-france-paris-landscape-975004/>
- <https://pixabay.com/vectors/toast-toaster-breakfast-kitchen-154616/>
- <https://pixabay.com/vectors/postage-stamp-postal-stamp-1479775/>
- <https://pixabay.com/vectors/iphone-cell-phone-phone-160307/>
- <https://pixabay.com/vectors/tag-luggage-label-blank-price-309129/>

## Sentiment analysis labels feelings in text



Sentiment Analysis identifies the emotional opinion of a writer's attitude. It is presented as a numerical score and a magnitude for the intensity of the feeling. It does not identify specific emotions but groups them into generally positive or negative or neutral, so "sad" and "angry" are both negative, and "funny" and "happy" are both positive. Because these are continuous values, you should define your own thresholds that work for your application. For example, maybe below 10% magnitude is not a strong enough feeling, or between -1.0 and +1.0 does not clearly express an emotion.

## Combined Entity and Sentiment Analysis



Entity Sentiment Analysis combines both entity analysis and sentiment analysis and attempts to determine the sentiment (positive or negative) expressed about entities within the text. Entity sentiment is represented by numerical score and magnitude values and is determined for each mention of an entity. Those scores are then aggregated into an overall sentiment score and magnitude for an entity.

<https://pixabay.com/vectors/pizza-food-slice-cheese-mushroom-23477/>  
<https://pixabay.com/vectors/essen-food-noodle-plate-spaghetti-1293911/>

## Content Classification



Sports



Foods



Literature

600+  
Categories



At the time of this writing, content is classified into 620 categories.

<https://cloud.google.com/natural-language/docs/categories>

<https://pixabay.com/photos/salad-fruits-berries-healthy-2756467/>

<https://pixabay.com/illustrations/games-sport-sports-competition-70833/>

<https://pixabay.com/photos/book-read-old-literature-pages-1659717/>



## Using the Natural Language API to Classify Unstructured Text

### Objectives

- Create a Natural Language API request
- Call the API with curl
- Use the NL API's text classification feature
- Use text classification to understand a dataset of news articles

Using the Natural Language API to classify unstructured text

<https://googlestaging.qwiklabs.com/labs/32334/>

## Module Summary

- Unstructured data is omnipresent... and really hard to make useful
- Fortunately, GCP has many APIs to make unstructured data useful

