**NOMBRE DEL PROYECTO**

**ANÁLISIS DE TEXTO BANCOLOMBIA**

**INTRODUCCIÓN:**

El principal objetivo de este análisis es realizar una comparación de ventajas y desventajas que ofrecen las APIS de Computer visión y Google Visión para extracción de texto, mas puntualmente para realizar un reconocimiento y extracción de caracteres de diferentes tipos de facturas.

**PROPUESTA 1**

**API DE SERVICIO COMPUTER VISION DE AZURE**

Extrae información muy completa de las imágenes para clasificar y procesar datos visuales y llevar a cabo una moderación automática de las imágenes.

Usa el reconocimiento óptico de caracteres (OCR) de última generación en la operación de lectura para detectar texto impreso y manuscrito insertado, extraer palabras reconocidas en secuencias de caracteres legibles por máquina y habilitar la búsqueda.

**Link:** <https://docs.microsoft.com/es-es/azure/cognitive-services/computer-vision/quickstarts/python-print-text>

**VENTAJAS**

Face Api es una de las características favoritas de esta aplicación, tiene un alto nivel de reconocimiento facial, el reconocimiento de personas famosas de todo el mundo es muy aceptable, lo que le permite trabajar y realizar todo tipo de pruebas con diferentes caras ya conocidas, además permite Reconocer todo tipo de expresiones faciales, estados de ánimo, sexo, dimensiones exactas de caras, sonrisas, también tiene la capacidad de detectar la edad de las caras a través de un algoritmo complejo, que realiza una operación matemática completa, basada en diferentes puntos de la cara, normalmente es muy preciso, es una herramienta muy completa.

Es una de las APIS más precisas para la visión por computadora, incluso la visión por computadora de Google no puede superarla, da la respuesta en un objeto JSON que podemos analizar fácilmente usando Python o cualquier otro lenguaje y la mejor parte es que su documentación es muy fácil para entender cualquier cuerpo y luego solo tienes que subir una imagen y obtendrás el resultado y se eliminarán todas las tareas tediosas como entrenar y probar el modelo.

El reconocimiento de la cara y la emoción es insuperable. Microsoft Vision API es mucho más avanzado en esta área que otro software de reconocimiento y en comparación con la API de Google también. Esta versión marca muy bien el contenido "para adultos". Proporciona algoritmos útiles que puedo usar para procesar imágenes específicas colocadas.

**Desventajas:**

El indexador de video, todavía está en una etapa beta, y no tiene todas las funciones para crear videos perfectos basados ​​en la compilación que se ha desarrollado en la API, esta es una desventaja con la que la aplicación cuenta y se espera que mejore pronto , para poder extraer todo tipo de expresiones de las caras y enviarlas a un video, para su posterior análisis.

Mientras se pasan fotogramas de una transmisión en vivo, los errores aumentan exponencialmente, solo pes posible proporcionar fotogramas estables como entrada.

OCR decepcionó un poco: parecía funcionar mejor en algunas imágenes y no tan bien con otras. Microsoft puede considerar renovar esta área para trabajar mejor con otros aspectos del software que parecen funcionar en un estándar más alto que el OCR.

Si va más allá del número mencionado de transacciones por segundo, su tiempo de respuesta se acelerará para estar dentro del límite acordado, por lo que si recibe ráfagas de llamadas, tendrá un problema.

El modelo de precios basado en el uso puede ser difícil de comprender e implementar controles de costos efectivos.

**CARACTERISTICAS:**

El precio del servicio de computer vision de cognitive services azure varia de acuerdo a estas características:

**-Caracteristica 1:** Etiqueta, Cara, GetThumbnail Color y Tipo de imagen

**-Caracteristica 2:** OCR, adulto, celebridad y punto de referencia

**-Característica 3:** Describir y reconocer texto

**link:** <https://azure.microsoft.com/es-es/services/cognitive-services/computer-vision/>

**OPERACIONES A REALIZAR CON COMPUTER VISION:**

**Tag**: Computer Vision API devuelve etiquetas basadas en más de 2,000 objetos reconocibles, seres vivos, paisajes y acciones. En los casos en que las etiquetas puedan resultar ambiguas o sean poco conocidas, la respuesta de la API ofrece "sugerencias" para aclarar el significado de la etiqueta.

**Face**: detecta rostros humanos dentro de una imagen.

**GetThumbnail**: después de cargar una imagen, GetThumbnail genera una miniatura de alta calidad. El algoritmo de Computer Vision API analiza los objetos dentro de la imagen y después la recorta para ajustarla a los requisitos de la región de interés.

**Color**: el algoritmo de Computer Vision extrae los colores de una imagen. Los colores se analizan en tres contextos diferentes: primer plano, segundo plano y completo. Los colores se agrupan en 12 colores de énfasis dominantes.

**Tipo de imagen**: Computer Vision API puede establecer una marca booleana para indicar si una imagen está en blanco y negro o en color, así como usar el mismo método para indicar si una imagen es un dibujo de línea o no. El tipo de imagen también indica si se trata de una imagen prediseñada o no y cuál es la calidad.

**OCR**: la tecnología del reconocimiento óptico de caracteres (OCR) detecta el contenido de texto de una imagen. El texto identificado se extrae en un flujo de caracteres de lectura mecánica para fines de búsqueda y muchos otros propósitos, que abarcan desde historias clínicas hasta seguridad y banca. Detecta automáticamente el idioma. OCR ahorra tiempo y proporciona comodidad a los usuarios, ya que les permite simplemente realizar fotos del texto en lugar de tener que transcribirlo. Consulte la [documentación](https://docs.microsoft.com/es-es/azure/cognitive-services/computer-vision/home) para conocer los idiomas compatibles.

**Adulto**: aplique la configuración para adultos para habilitar la restricción automatizada de contenido para adultos en imágenes.

**Celebridad**: el modelo de reconocimiento de celebridades de Azure reconoce 200,000 celebridades de los negocios, la política, el deporte y el espectáculo de todo el mundo.

**Analizar**: invoque varias operaciones al mismo tiempo. Especifique qué funciones desea ejecutar, y la API las ejecutará todas juntas. Cada operación incluida en la operación de "análisis" se contará como una transacción independiente.

**Costos de API**

**Basados en el precio incluyendo las 3 caracteristicas principales:**

**1.**Etiqueta, Cara, GetThumbnail Color y Tipo de imagen

**2.** OCR, adulto, celebridad y punto de referencia

**3.** Describir y reconocer texto.

|  |  |  |  |
| --- | --- | --- | --- |
| **Instancia** | **Numero Transacciones** | **Precio Mensual** | |
| Gratis | 5000(20 por minuto) | $0 usuarios Nuevos | |
| S1 | 10000(hasta 10 x Segundo) | Característica 1: | $10 US |
| Característica 2: | $15 US |
| Característica 3: | $25 US |
| TOTAL | $ 100 US |
| S1 | 20000(hasta 10 x Segundo) | Característica 1: | $20 US |
| Característica 2: | $30 US |
| Característica 3: | $50 US |
| TOTAL | $ 100 US |

La tabla muestra el ejemplo de costos hasta para 20000 transacciones, sin embargo en la calculadora de precios se puede escoger numero que mas se acomode a las necesidades.

**Link Calculadora de precios:**

**[https://azure.microsoft.com/es-es/pricing/calculator/#cognitive-services3f11bc83-e3e7-4f08-bb1b-872319472b66](https://azure.microsoft.com/es-es/pricing/calculator/" \l "cognitive-services3f11bc83-e3e7-4f08-bb1b-872319472b66)**

**Que es el Numero de Transacciones:**

Para el reconocimiento de texto, cada llamada POST cuenta como una transacción. Todas las llamadas GET para ver los resultados del servicio asincrónico se cuentan como una transacción, pero son gratuitas. Para todas las demás operaciones, cada llamada de característica cuenta como una transacción, ya sea que se llame independientemente o se agrupe a través de la llamada Analizar. Las llamadas Analizar se utilizan para facilitar la llamada a la API, pero cada característica utilizada cuenta como una transacción. Por ejemplo, una llamada Analizar que contenga Etiqueta, Cara y Adulto contará como tres transacciones.

Consulte la [documentación](https://docs.microsoft.com/es-es/azure/cognitive-services/computer-vision/home) para una lista completa y descripciones detalladas de las operaciones.

**Llamada post:** POST consiste en datos "ocultos" (porque el cliente no los ve) enviados por un formulario cuyo método de envío es post. Es adecuado para formularios. Los datos no son visibles.

**Llamada post:** GET lleva los datos de forma "visible" al cliente (navegador web). El medio de envío es la URL. Los datos los puede ver cualquiera.

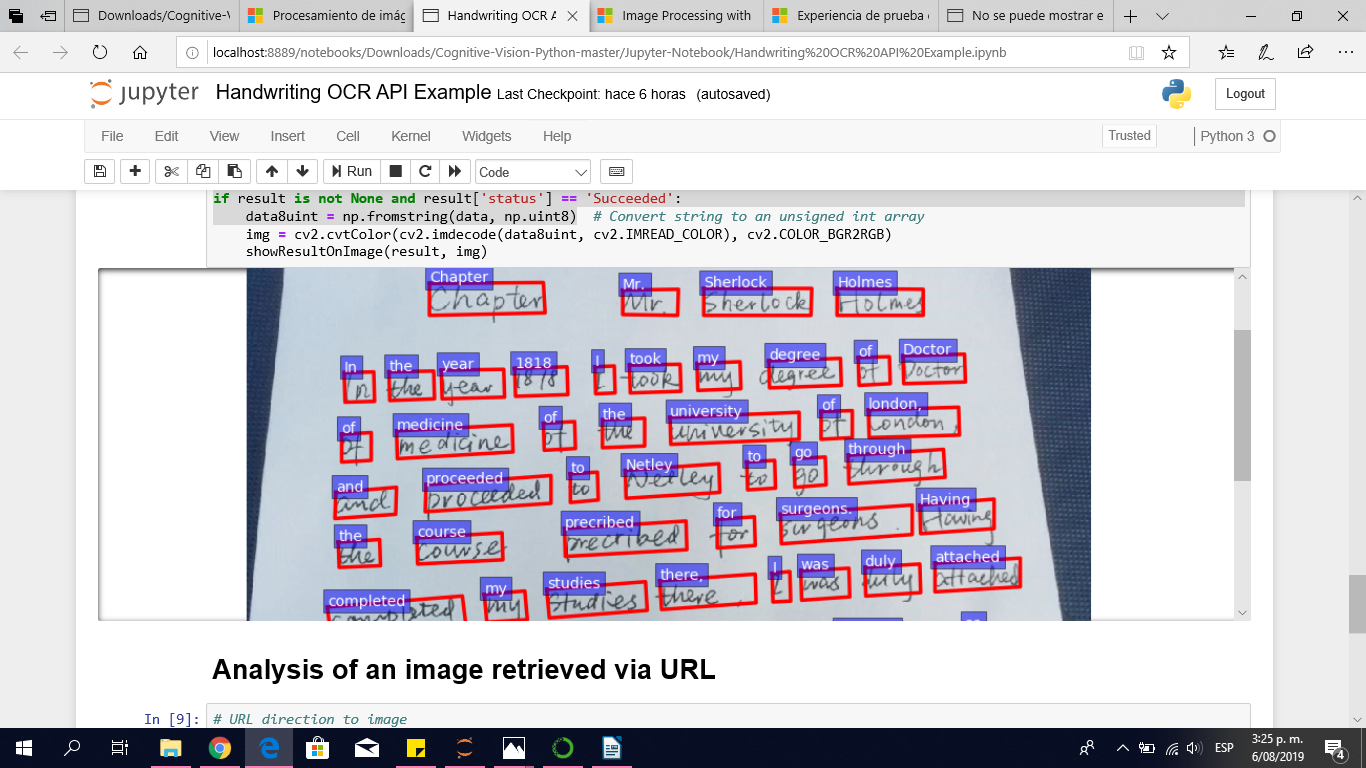
**Link:** [**https://azure.microsoft.com/es-es/pricing/details/cognitive-services/computer-vision/**](https://azure.microsoft.com/es-es/pricing/details/cognitive-services/computer-vision/)

**Computer vision contiene el Programa de licencias Contrato de Microsoft Online Services**

Este es el plan de pago más popular y flexible. Con los precios de pago por uso de Azure, no hay requisitos de compra mínima ni compromisos. Se uede pagar con tarjeta de crédito o mediante factura y cancelar la suscripción en cualquier momento

Link: **[https://azure.microsoft.com/es-es/pricing/calculator/#cognitive-services263f296d-a730-492e-82b1-0bf66094d528](https://azure.microsoft.com/es-es/pricing/calculator/" \l "cognitive-services263f296d-a730-492e-82b1-0bf66094d528)**

**Vista de la API:**



**Extraer varias imágenes:**

La idea es que atraves de la API de Computer Vision Sea posible procesar múltiples imágenes, la url a continuación muestra información de como se haría dicha implementación:

**Link:** [**https://www.jesusninoc.com/01/08/extraer-datos-de-varias-imagenes-mediante-el-reconocimiento-optico-de-caracteres-y-el-analisis-de-imagenes-computer-vision-api-de-microsoft-azure/**](https://www.jesusninoc.com/01/08/extraer-datos-de-varias-imagenes-mediante-el-reconocimiento-optico-de-caracteres-y-el-analisis-de-imagenes-computer-vision-api-de-microsoft-azure/)

**PROPUESTA 2**

**1. Google Vision:**

Extrae información valiosa de tus imágenes en la nube o en el perímetro con AutoML Vision, o usa modelos previamente preparados de la API Vision para detectar emociones, interpretar texto y mucho más**.**

La API Vision de Google Cloud ofrece modelos de aprendizaje automático previamente preparados y muy potentes a través de las API REST y RPC. Asigna etiquetas a imágenes y clasifícalas rápidamente en millones de categorías predefinidas. Detecta objetos y caras, lee texto impreso y manuscrito, y consigue metadatos de gran valor para tu catálogo de imágenes.

**VENTAJAS:**

#### **Detecta objetos automáticamente**

Detecta y clasifica distintos objetos, y consigue información como la posición de cada uno de ellos dentro de la imagen. Consulta más información sobre la detección de objetos con la [API Vision](https://cloud.google.com/vision/docs/object-localizer/?hl=es) y [AutoML Vision](https://cloud.google.com/vision/automl/object-detection/docs/?hl=es)

#### **Comprende el texto y actúa en consecuencia**

La API Vision utiliza la tecnología de reconocimiento óptico de caracteres (OCR) para [detectar texto](https://cloud.google.com/vision/docs/detecting-fulltext/?hl=es) en imágenes en más de 50 [idiomas](https://cloud.google.com/vision/docs/languages/?hl=es) y diferentes [tipos de archivo](https://cloud.google.com/vision/docs/supported-files/?hl=es). También forma parte de [IA para interpretar documentos](https://cloud.google.com/solutions/document-understanding/?hl=es), que permite procesar rápidamente millones de documentos y automatizar los flujos de trabajo de las empresas.

No tienen un costo inicial y es un servicio de pago por uso sin cargos por cancelación.

Google viene con el más alto control de seguridad, privacidad y cumplimiento donde proporciona la certificación ISO y SOC para los productos

Cloud API de visión puede detectar cualquier objeto usando Google Image Search.Cloud Vision puede moderar la búsqueda de contenido utilizando el poderoso Google Analytics.

Se pueden aplicar múltiples parámetros de filtro a una sola imagen.

**DESVENTAJAS:**

Cloud Datastore tiene características limitadas y es difícil cambiarlo

Google CDN no funciona con fuentes personalizadas

StackDriver (registros de GCP) se siente separado

El reconocimiento facial no es compatible.

**LINK:** [**https://blog.api.rakuten.net/top-machine-vision-apis/**](https://blog.api.rakuten.net/top-machine-vision-apis/)

**COSTOS DE API:**

| PRECIO POR 1000 UNIDADES (USO MENSUAL) | | |  |
| --- | --- | --- | --- |
| **FUNCIÓN** | **1-1000 unidades al mes** | **1001-5.000.000 unidades al mes** | **5.000.001-20.000.000 unidades al mes** |
| Detección de etiquetas | Gratis | 1,50 USD | 1 USD |
| Detección de texto | Gratis | 1,50 USD | 0,60 USD |
| Detección de Búsqueda Segura (contenido explícito) | Gratis | Gratis con la detección de etiquetas, o bien 1,50 USD | Gratis con la detección de etiquetas, o bien 0,60 USD |
| Detección facial | Gratis | 1,50 USD | 0,60 USD |
| Detección de puntos de referencia | Gratis | 1,50 USD | 0,60 USD |
| Detección de logotipos | Gratis | 1,50 USD | 0,60 USD |
| Propiedades de imagen | Gratis | 1,50 USD | 0,60 USD |
| Pistas de recorte | Gratis | Gratis con propiedades de la imagen, o bien 1,50 USD | Gratis con propiedades de la imagen, o bien 0,60 USD |
| Detección web | Gratis | 3,50 USD | n/a |
| Detección de texto en documentos | Gratis | 1,50 USD | 0,60 USD |
| Localizador de objetos | Gratis | 2,25 USD | 1,50 USD |

**Numero de Unidades:**

Los cargos se aplican por imagen. En el caso de los archivos con varias páginas, como los PDF, cada página se trata como una imagen independiente.

Cada función que se aplica a una imagen se considera una *unidad* facturable. Por ejemplo, si aplicas la detección de caras y la de etiquetas a la misma imagen, se te facturará por 1 unidad de detección de caras y 1 unidad de detección de etiquetas.

## **Ejemplo**

Si tu aplicación realizó las siguientes solicitudes a lo largo de un mes:

* 700 imágenes con detección de etiquetas
* 5300 imágenes con detección de puntos de referencia

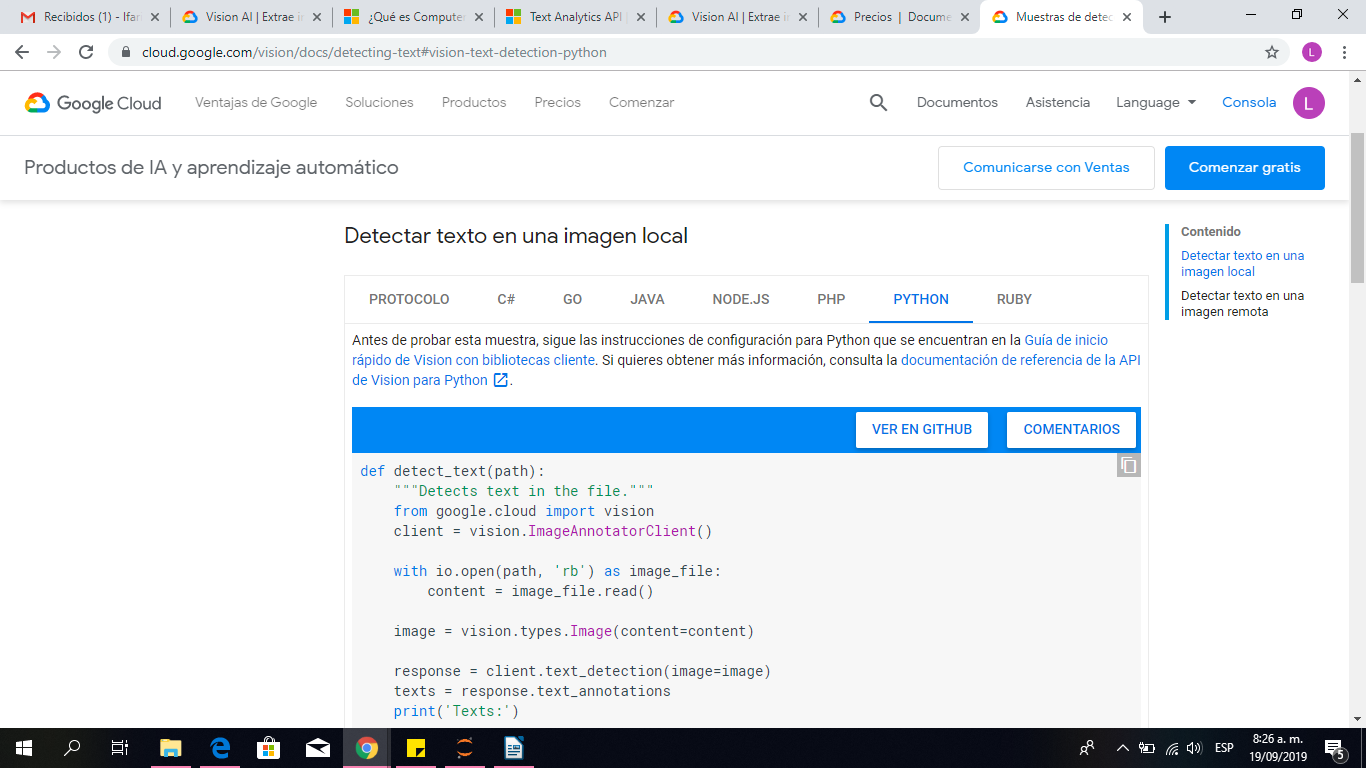
El precio sería el siguiente:

* 0 USD por las 700 solicitudes de detección de etiquetas.
* 0 USD por las primeras 1000 solicitudes de detección de puntos de referencia.
* 7,50 USD por las 4300 solicitudes de detección de puntos de referencia restantes. Como el precio se calcula por cada 1000 unidades, el de 4300 solicitudes sería 5 x 1,50 USD.

El coste total ascendería a 7,50 USD.

**Link:** [**https://cloud.google.com/vision/pricing**](https://cloud.google.com/vision/pricing)

**VISTA DE API:**

****

**Comparación de Herramientas:**

Se puede visualizar en el siguiente link: <https://blog.api.rakuten.net/top-machine-vision-apis/>