

OCR - Android



UNIVERSIDAD DE GRANADA

Periféricos y Dispositivos de Interfaz Humana

Rafael Luque Pérez-Vico

1. ¿Qué es el OCR?

Según Jesús Tomas en su obra "Dispositivos Wearables, Visión Artificial, Google Glass y Android TV", OCR (Reconocimiento Óptico de Caracteres) se define como "los sistemas que, a partir de un texto escrito o impreso en papel u otro soporte similar, generan un archivo de texto en un medio de almacenamiento informático".

En otras palabras, OCR es una tecnología que permite convertir imágenes y documentos escritos a mano en texto digital editables. Esta tecnología es capaz de reconocer caracteres, números y símbolos en documentos escaneados o imágenes, y luego los convierte en texto que se puede buscar, editar y almacenar electrónicamente. En esencia, el OCR actúa como un puente entre el mundo físico y el mundo digital, permitiendo la automatización de tareas que antes eran laboriosas y propensas a errores.

2. Proceso OCR

Las computadoras suelen enfrentar dificultades para detectar y reconocer texto en imágenes, especialmente cuando se trata de caracteres con diferentes estilos y fuentes. Este desafío se agrava aún más cuando se trata de texto manuscrito. A pesar de esto, los algoritmos de reconocimiento óptico de caracteres (OCR) abordan este problema de manera eficiente. Estos algoritmos trabajan mediante un proceso que convierte imágenes de texto en formato textual, permitiendo así que las computadoras puedan interpretar y procesar dicho texto, incluso cuando está escrito a mano.

A continuación veremos un paso a paso sobre cómo OCR trabaja :

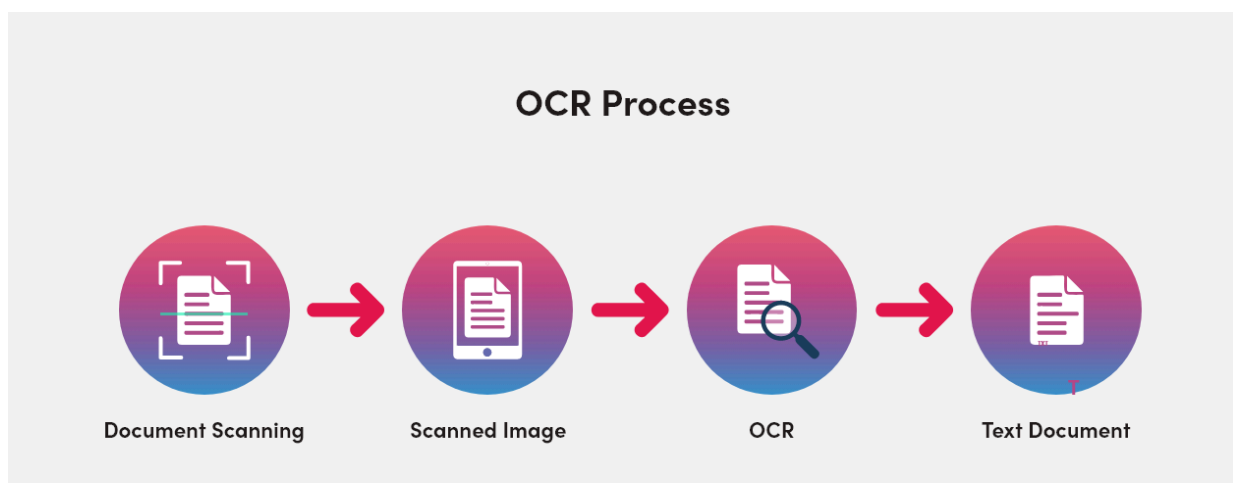


Figura 1 : Paso a paso de un proceso OCR

- Escaneado de Documentos

Durante la primera etapa del proceso, se utilizan escáneres ópticos para digitalizar documentos. Estos dispositivos limpian la imagen y separan los caracteres del resto del contenido. Una vez completado el escaneo, la imagen se guarda en un mapa de bits.

- Refinado de Imágenes

En el siguiente paso, el software de OCR refina el texto al alinear perfectamente y suavizar sus píxeles haciendo que el texto sea claro y legible. Además, puedes aplicar técnicas de procesamiento de imágenes para mejorar o iluminar las imágenes, de modo que los programas de OCR pueden leer el texto de manera óptima.

- Segmentación

Tras el proceso de refinado de imagen, procedemos a dividir los caracteres del texto en grupos similares y significativos. Una técnica de segmentación de imágenes divide todo el texto en colores blanco y negro. Para distinguir el texto del fondo, se identifica el área negra como caracteres, mientras que el área blanca se considera fondo, lo que facilita el reconocimiento óptico de caracteres.

- Reconocimiento de Caracteres

En este punto, los algoritmos de OCR realizan diversas manipulaciones para identificar letras y números en negro. Normalmente, OCR aprovecha dos enfoques para el reconocimiento óptico de caracteres: el reconocimiento de patrones y la detección de características.

- Reconocimiento de Patrones

Debido a que los algoritmos de OCR están diseñados para ajustarse a una variedad de fuentes, formatos de texto y estilos de escritura, son capaces de distinguir los caracteres del texto escaneado. Este enfoque habilita al software de OCR para identificar el patrón específico del texto.

- Detección de Características

Esta técnica, también llamada extracción de características o reconocimiento inteligente de caracteres (ICR), identifica las características del texto de una manera considerablemente más sofisticada de lo que podríamos anticipar. Por ejemplo, puede detectar rasgos y patrones como líneas, bucles cerrados, curvas,

dirección de líneas e intersecciones. Luego, el OCR encuentra la mejor coincidencia para estas características entre las letras correspondientes.

- Post-Procesamiento

Usualmente, las salidas de OCR se presentan como una cadena de caracteres o un archivo. Después de que el reconocimiento de caracteres se complete exitosamente, es importante verificar los resultados. Aunque no hay herramientas dedicadas que garanticen una precisión del 100%, el software de OCR suele ofrecer una precisión del 98% al 99%. Sin embargo, los textos escritos a mano de manera deficiente y los patrones de texto extraños pueden aumentar la tasa de error hasta un 20%.

Por lo general, los resultados se verifican utilizando el diccionario interno del software de OCR. Se compara el último resultado de OCR con la versión original del contenido.

Existen dos formas de medir las salidas del OCR:

- Precisión a nivel de caracteres

Muestra cuántos caracteres son detectados correctamente.

- Precisión a nivel de palabras

Indica cuántas palabras se identifican correctamente.

3.API/Bibliotecas OCR

- ML Kit de Firebase - Reconocimiento de Texto v2

Firebase Machine Learning es una plataforma que ofrece herramientas y servicios diseñados para integrar el aprendizaje automático en aplicaciones móviles, ya sea en sistemas Android o IOS. Entre sus componentes, incluye un conjunto de API en la nube que permite realizar diversas tareas, como reconocimiento de texto, identificación de puntos de referencia, etiquetado de imágenes, etc.

En concreto, nosotros usaremos la API de reconocimiento de texto en imágenes del Kit de AA v2 que puede reconocer texto en cualquier grupo de caracteres. Las funciones clave de esta API son :

- Admite el reconocimiento de texto en alfabetos chinos, japonés latinos
- Admite la detección de símbolos, elementos, líneas y párrafos
- Identifica el idioma del texto reconocido
- Puede reconocer texto en tiempo real en una amplia variedad de dispositivos

El reconocedor de texto segmenta el texto en bloques, líneas, elementos y símbolos. A grandes rasgos:

- Un bloque es una agrupación consecutiva de líneas de texto, como un párrafo o una columna
- Una línea es una secuencia consecutiva de palabras en la misma línea horizontal
- Un elemento es un conjunto consecutivo de caracteres alfanuméricos en la misma línea horizontal
- Un símbolo es un carácter alfanumérico único en la misma línea horizontal, en la mayoría de los idiomas latios

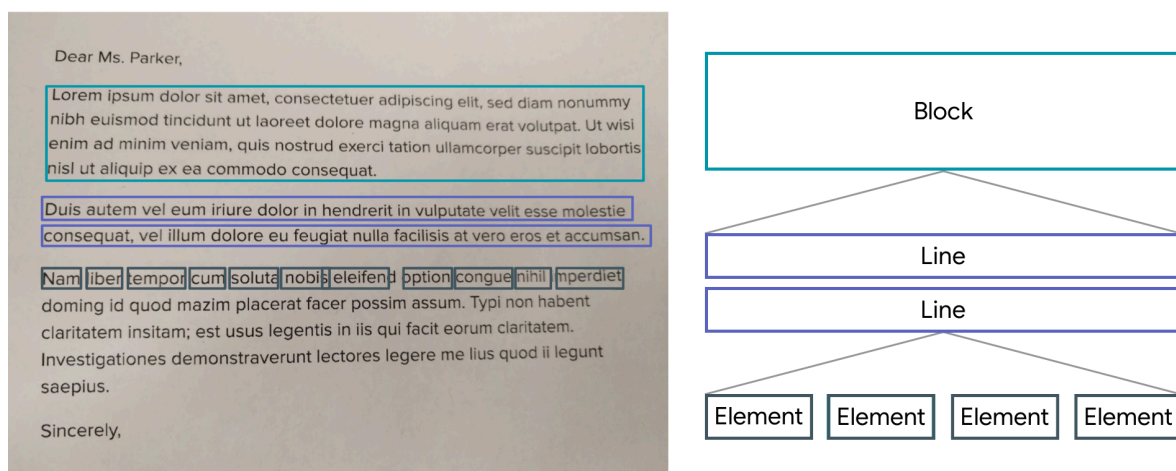


Figura 2 : Ejemplo de la API de Reconocimiento de Texto v2

</> Reconocimiento de Texto V2

- Tesseract

La biblioteca Tesseract de OCR es una herramienta de código abierto ampliamente reconocida y utilizada para el reconocimiento óptico de caracteres. Desarrollada originalmente en HP Labs en la década de 1980 y posteriormente mantenida por Google, Tesseract ha evolucionado para convertirse en una solución poderosa y versátil para la extracción de texto a partir de imágenes y documentos escaneados.

La funcionalidad principal de Tesseract radica en su capacidad para procesar imágenes de texto y convertirlas en texto digital editable. Utiliza algoritmos avanzados de visión por computadora y aprendizaje automático para identificar patrones de texto en las imágenes, segmentar caracteres y palabras, y finalmente reconocer y transcribir el texto con alta precisión.

Una de las características distintivas de Tesseract es su capacidad para reconocer una amplia gama de idiomas y tipos de fuente, lo que lo hace adecuado para aplicaciones globales y multilingües. Además, Tesseract es altamente adaptable y puede ser integrado en una variedad de plataformas y entornos de desarrollo, incluyendo sistemas operativos como Windows, macOS y Linux, así como también en aplicaciones web y móviles.

La precisión y el rendimiento de Tesseract pueden mejorarse aún más mediante técnicas de preprocesamiento de imágenes, como el ajuste de contraste, la corrección de distorsiones y el filtrado de ruido. Además, Tesseract ofrece opciones de configuración avanzadas que permiten ajustar parámetros específicos para adaptarse mejor a diferentes tipos de documentos y condiciones de escaneo.

Otro aspecto destacado de Tesseract es su activa comunidad de desarrollo y soporte, que continúa mejorando y actualizando la biblioteca con nuevas características y mejoras. Esto garantiza que Tesseract siga siendo una opción líder para aplicaciones de OCR en una amplia variedad de industrias y casos de uso, desde la digitalización de documentos hasta la automatización de procesos empresariales.

</> Github Tesseract-OCR

- Android Image Cropper

La biblioteca Android Image Cropper es una potente herramienta que facilita el recorte de imágenes en aplicaciones Android. Permite a los usuarios seleccionar una imagen de su galería o tomar una foto con la cámara y luego editarla para hacerla según sus necesidades.

Algunas de las funcionalidades que incluye son:

- Selección de imagen desde la galería o tomar una nueva con la cámara en ese instante
- Proporciona una interfaz de usuario intuitiva para recortar la imagen seleccionada
- Permite ajustar aspectos como el tamaño del recorte, la relación de aspecto y la calidad de imagen
- Ofrece manejo de errores y excepciones para garantizar una experiencia de usuario fluida

- Es compatible con diferentes versiones de Android y es fácil de integrar en aplicaciones existentes

[`</>` Github Android Image Cropper](#)

4. Casos de Uso de la Tecnología OCR

El reconocimiento óptico de caracteres (OCR) se encuentra ampliamente integrado en numerosos aspectos de nuestra vida diaria, abarcando una diversidad de sectores e aplicaciones tales como:

Banca

En el ámbito bancario, el OCR se emplea para agilizar y asegurar la verificación de documentación relacionada con préstamos, depósitos y otras operaciones financieras. Esta aplicación ha fortalecido notablemente la prevención de fraudes y la seguridad en las transacciones. Por ejemplo, *BlueVine*, una empresa de tecnología financiera que proporciona financiamiento a pequeñas y medianas empresas, recurrió a Amazon Textract, un servicio de OCR basado en la nube, para desarrollar una solución que permitiera a las pequeñas empresas en los Estados Unidos acceder rápidamente a préstamos del Programa de Protección de Cheques de Pago (PPP) como parte del paquete de estímulo por la crisis de COVID-19. Amazon Textract procesó y analizó automáticamente decenas de miles de formularios PPP al día, lo que habilitó a *BlueVine* a asistir a miles de empresas en la obtención de fondos, preservando así más de 400,000 puestos de trabajo en el proceso.

Sanidad

En el ámbito de la salud, el OCR desempeña un papel fundamental en el procesamiento de los registros de pacientes, abarcando tratamientos, pruebas, historiales hospitalarios y transacciones de seguros. Esta tecnología contribuye a optimizar los flujos de trabajo y a reducir la carga de trabajo manual en los hospitales, al tiempo que garantiza la actualización constante de los registros. Por ejemplo, nib Group, que brinda seguros médicos y de salud a más de un millón de australianos, enfrenta diariamente un gran volumen de solicitudes médicas. Sus clientes tienen la opción de capturar fotos de sus facturas médicas y enviarlas a través de la aplicación móvil de nib. Gracias a Amazon Textract, estas imágenes se procesan

automáticamente, lo que permite que la empresa apruebe los reclamos de manera mucho más rápida y eficiente.

Logística

Las empresas de logística aprovechan el OCR para mejorar la eficiencia en el seguimiento de etiquetas de paquetes, facturas, recibos y otros documentos. Por ejemplo, Foresight Group ha implementado Amazon Textract para automatizar el procesamiento de facturas en SAP. Anteriormente, la introducción manual de estos documentos comerciales era una tarea laboriosa y propensa a errores, ya que los empleados de Foresight tenían que ingresar los datos en varios sistemas de contabilidad. Con Amazon Textract, el software de Foresight puede reconocer caracteres con mayor precisión en una variedad de diseños, lo que se traduce en una mejora significativa en la eficiencia operativa.

5. Bibliografía

- Dar, T. (2022, 18 noviembre). *Optical Character Recognition (OCR) — A complete guide*. Revolveai. <https://revolveai.com/optical-character-recognition/>
- Manuel, A. P. J., & De Valladolid Escuela Técnica Superior de Ingenieros de Telecomunicación, U. (2018). *Diseño y desarrollo de una aplicación Android para el reconocimiento óptico de caracteres*. Universidad de Valladolid. <https://uvadoc.uva.es/handle/10324/32987>
- Bustos, J. L. (2024, 24 abril). Qué es el ML Kit de Firebase | KeepCoding Bootcamps. *KeepCoding Bootcamps*. https://keepcoding.io/blog/que-es-el-ml-kit-de-firebase/#Que_es_el_ML_Kit_de_Firebase
- ArthurHub. (s. f.). *GitHub - ArthurHub/Android-Image-Cropper: Image Cropping Library for Android, optimized for Camera / Gallery*. GitHub. <https://github.com/ArthurHub/Android-Image-Cropper?tab=readme-ov-file>

- Tesseract-Ocr. (s. f.). *GitHub - tesseract-ocr/tesseract: Tesseract Open Source OCR Engine (main repository)*. GitHub.
<https://github.com/tesseract-ocr/tesseract?tab=readme-ov-file>
- *Reconocimiento de texto v2*. (s. f.). Google For Developers.
<https://developers.google.com/ml-kit/vision/text-recognition/v2?hl=es-419>
- *¿Qué es el OCR? - Explicación del reconocimiento óptico de caracteres - AWS*. (s. f.). Amazon Web Services, Inc. **<https://aws.amazon.com/es/what-is/ocr/>**