

DIGITALIZACIÓN

1. TECNOLOGÍAS OCR (RECONOCIMIENTO ÓPTICO DE CARACTERES)

1.1 Fundamentos del OCR

El Reconocimiento Óptico de Caracteres (OCR) es una tecnología que permite convertir diferentes tipos de documentos (como imágenes escaneadas, archivos PDF o fotografías) en datos editables y con capacidad de búsqueda.

Proceso técnico del OCR:

1. Preprocesamiento de la imagen:

- Corrección de rotación y alineación
- Eliminación de ruido
- Normalización de contraste y brillo
- Binarización (conversión a blanco y negro)

2. Segmentación:

- División del documento en zonas (textos, imágenes, tablas)
- División de texto en líneas
- División de líneas en palabras
- División de palabras en caracteres individuales

3. Reconocimiento de caracteres:

- Extracción de características (trazos, contornos, intersecciones)
- Comparación con patrones predefinidos
- Uso de redes neuronales o algoritmos de machine learning
- Análisis contextual (diccionarios, gramática)

4. Post-procesamiento:

- Corrección ortográfica
- Reconstrucción de formato
- Generación del documento final

1.2 Tecnologías avanzadas de OCR

OCR con Inteligencia Artificial

- **Redes neuronales convolucionales (CNN):** Mejoran significativamente la precisión del reconocimiento.
- **Aprendizaje profundo (Deep Learning):** Permite reconocer textos en condiciones complejas (manuscritos, documentos deteriorados).
- **Aprendizaje por refuerzo:** Sistemas que mejoran con el uso continuo.

Reconocimiento de escritura manuscrita (ICR - Intelligent Character Recognition)

- Tecnología especializada en interpretar texto escrito a mano.
- Aplicaciones en formularios, notas, documentos históricos.
- Desafíos: variabilidad de estilo, caligrafías personales, calidad del trazo.

Reconocimiento de campos estructurados

- **OMR (Optical Mark Recognition):** Reconocimiento de marcas como casillas marcadas en formularios.
- **BCR (Bar Code Recognition):** Lectura de códigos de barras 1D y 2D.
- **Reconocimiento de formularios:** Extracción automática de datos de campos predefinidos.

1.3 Aplicaciones prácticas del OCR en entornos empresariales

Automatización de procesos documentales

- **Procesamiento de facturas:** Extracción automática de datos relevantes (número de factura, importes, fechas).
- **Clasificación automática de documentos:** Identificación del tipo de documento para su enrutamiento.
- **Indexación automática:** Generación de metadatos para búsqueda y recuperación eficiente.

Optimización de flujos de trabajo

- Integración con sistemas ERP y CRM
- Automatización de la entrada de datos
- Reducción de tiempos de procesamiento y errores humanos

OCR en movilidad

- Aplicaciones móviles para captura de documentos
- Procesamiento en tiempo real
- Uso en auditorías, inspecciones o trabajo de campo

1.4 Herramientas y tecnologías OCR

Software OCR comercial

- **ABBYY FineReader:** Solución profesional de alta precisión.
- **Adobe Acrobat Pro DC:** Capacidades OCR integradas.
- **OmniPage:** Especializado en reconocimiento avanzado y conversión de formatos.
- **IRIS PowerScan:** Enfocado en volúmenes altos de documentación.

Soluciones OCR de código abierto

- **Tesseract OCR:** Desarrollado por Google, considerado uno de los motores OCR libres más precisos.
- **OCRmyPDF:** Herramienta para añadir capas de OCR a documentos PDF.
- **OCRopus:** Framework modular para procesamiento de documentos.
- **Kraken:** Especializado en textos históricos.

APIs y servicios en la nube

- **Google Cloud Vision API:** Reconocimiento avanzado basado en IA.
- **Microsoft Azure Computer Vision:** Reconocimiento de texto integrado con otros servicios cognitivos.
- **Amazon Textract:** Extracción de texto y datos estructurados de documentos.

2. TECNOLOGÍA HRT (HUMAN RECOGNITION TECHNOLOGY)

2.1 Fundamentos de HRT

Las tecnologías de reconocimiento humano (HRT) engloban diversos sistemas que identifican o verifican la identidad de una persona a través de características físicas, comportamientos o una combinación de ambos.

2.2 Principales tecnologías biométricas

Biometría física

- **Reconocimiento facial:** Análisis de características faciales y sus relaciones espaciales.
- **Huella dactilar:** Lectura y comparación de los patrones únicos de las huellas digitales.
- **Escaneo de iris/retina:** Análisis de los patrones del iris o la retina del ojo.
- **Reconocimiento de voz:** Análisis de patrones de voz y fonación.
- **Geometría de la mano:** Medición de la forma y tamaño de la mano.
- **Reconocimiento vascular:** Análisis del patrón de venas en dedos o palma.

Biometría comportamental

- **Dinámica de tecleo:** Análisis del ritmo y patrón al escribir en un teclado.
- **Firma dinámica:** Análisis no sólo de la forma de la firma sino de la presión, velocidad y ritmo al firmar.
- **Análisis de marcha:** Reconocimiento basado en la forma de caminar.
- **Patrones de navegación:** Comportamiento del usuario en entornos digitales.

2.3 Aplicaciones en entornos empresariales y de seguridad

Control de acceso físico y lógico

- Acceso a instalaciones mediante reconocimiento facial o huella dactilar.
- Inicio de sesión en sistemas informáticos mediante biometría.
- Acceso a zonas restringidas con múltiples factores biométricos.

Identificación y verificación

- KYC (Know Your Customer) en sectores financieros y seguros.
- Verificación de identidad para acceso remoto a servicios.
- Autenticación en transacciones de alto valor.

Seguridad documental

- Pasaportes y documentos de identidad electrónicos.
- Certificados y firmas digitales con componente biométrico.
- Autenticación de documentos sensibles.

2.4 Consideraciones sobre privacidad y seguridad

- **Protección de datos biométricos:** Encriptación y almacenamiento seguro.
- **Normativa RGPD:** Consideraciones específicas para datos biométricos.
- **Consentimiento explícito:** Requisitos legales para la recogida y uso.
- **Prevención del spoofing:** Mecanismos contra suplantación de identidad.

3. CLOUD COMPUTING EN DIGITALIZACIÓN

3.1 Fundamentos del Cloud Computing

El cloud computing o computación en la nube se refiere al suministro de recursos informáticos (servidores, almacenamiento, bases de datos, redes, software) a través de Internet, ofreciendo flexibilidad, escalabilidad y modelo de pago por uso.

3.2 Modelos de servicio en la nube

IaaS (Infraestructura como Servicio)

- **Definición:** Provisión de recursos de infraestructura virtualizados a través de la red.
- **Componentes:** Servidores virtuales, almacenamiento, redes y firewalls.
- **Ventajas:** Control total sobre los sistemas operativos y aplicaciones.
- **Ejemplos:** Amazon EC2, Microsoft Azure VMs, Google Compute Engine.
- **Aplicación en digitalización:** Infraestructura flexible para sistemas de gestión documental.

PaaS (Plataforma como Servicio)

- **Definición:** Entorno de desarrollo e implementación completo en la nube.
- **Componentes:** Sistemas operativos, middleware, herramientas de desarrollo.
- **Ventajas:** Desarrollo rápido de aplicaciones sin gestionar infraestructura.
- **Ejemplos:** Azure App Service, Google App Engine, AWS Elastic Beanstalk.
- **Aplicación en digitalización:** Desarrollo de aplicaciones de procesamiento documental.

SaaS (Software como Servicio)

- **Definición:** Aplicaciones completas entregadas a través de Internet.
- **Componentes:** Aplicaciones empresariales accesibles vía web.
- **Ventajas:** Sin instalación local, actualizaciones automáticas.
- **Ejemplos:** Google Workspace, Microsoft 365, Salesforce.
- **Aplicación en digitalización:** Servicios de gestión documental en la nube.

3.3 Infraestructura cloud para gestión documental

Almacenamiento en la nube para documentos

- **Almacenamiento de objetos:** Amazon S3, Azure Blob Storage, Google Cloud Storage.

- **Sistemas de archivos en la nube:** Azure Files, Amazon EFS.
- **Consideraciones de diseño:** Redundancia, replicación geográfica, ciclo de vida.

Bases de datos para metadatos documentales

- **SQL:** Azure SQL, Amazon RDS, Google Cloud SQL.
- **NoSQL:** MongoDB Atlas, DynamoDB, Cosmos DB.
- **Bases de datos de búsqueda:** Elasticsearch, Algolia.

Servicios cognitivos en la nube

- **Procesamiento de lenguaje natural:** Comprensión de documentos.
- **Servicios de visión artificial:** OCR avanzado en la nube.
- **Machine Learning:** Clasificación automática de documentos.

3.4 Ventajas del cloud en procesos de digitalización

Escalabilidad y elasticidad

- Adaptación a picos de procesamiento documental.
- Crecimiento del almacenamiento según necesidades.
- Pago por uso.

Accesibilidad y trabajo colaborativo

- Acceso desde cualquier ubicación y dispositivo.
- Colaboración en tiempo real sobre documentos.
- Integración con herramientas de workflow.

Seguridad y cumplimiento normativo

- Certificaciones internacionales (ISO 27001, SOC 2).
- Cifrado end-to-end.
- Controles de acceso granulares.
- Auditoría y trazabilidad.

3.5 Tendencias en cloud para digitalización

- **Serverless computing:** Procesamiento documental sin gestionar servidores.
- **Edge computing:** Procesamiento cercano al punto de captura.
- **Multi-cloud e Hybrid-cloud:** Estrategias combinadas para optimizar recursos.
- **Contenedores:** Despliegue ágil de aplicaciones de procesamiento documental.

4. VIRTUALIZACIÓN EN ENTORNOS DE DIGITALIZACIÓN

4.1 Fundamentos de la virtualización

La virtualización es la creación de versiones virtuales (no físicas) de recursos como servidores, dispositivos de almacenamiento, redes y sistemas operativos, permitiendo un uso más eficiente del hardware.

4.2 Tipos de virtualización

Virtualización de servidores

- **Hipervisor Tipo 1 (bare-metal):** VMware ESXi, Microsoft Hyper-V, KVM.
- **Hipervisor Tipo 2 (hosted):** VMware Workstation, Oracle VirtualBox.
- **Beneficios para sistemas de digitalización:** Consolidación de servidores, aislamiento de aplicaciones, mejor aprovechamiento de recursos.

Virtualización de almacenamiento

- **SAN virtuales:** Abstracción del almacenamiento físico.
- **Software-Defined Storage:** Gestión centralizada del almacenamiento.
- **Aplicación en digitalización:** Repositorios documentales flexibles y escalables.

Virtualización de redes (SDN)

- **Redes definidas por software:** Separación del control y plano de datos.
- **Microsegmentación:** Políticas de seguridad granulares.
- **Aplicación en digitalización:** Gestión segura del tráfico documental.

Virtualización de escritorios (VDI)

- **Definición:** Separación del entorno de escritorio del dispositivo físico.
- **Implementaciones:** Citrix Virtual Apps and Desktops, VMware Horizon.
- **Aplicación en digitalización:** Puestos de digitalización virtualizados, acceso seguro a aplicaciones de procesamiento.

4.3 Containers y microservicios

Tecnologías de contenedores

- **Docker:** Plataforma para desarrollar y ejecutar aplicaciones en contenedores.
- **Kubernetes:** Orquestación de contenedores para despliegues complejos.
- **Docker Swarm:** Orquestación nativa de Docker.

Arquitectura de microservicios

- **Definición:** Aplicaciones como conjunto de servicios independientes.
- **Ventajas:** Desarrollo ágil, escalabilidad, resiliencia.
- **Aplicación en digitalización:** Descomposición de procesos documentales en microservicios.

Contenedores en procesos de digitalización

- Servicios de OCR containerizados.
- Procesamiento distribuido de documentos.
- Despliegue ágil de nuevas funcionalidades.

4.4 Infraestructura hiperconvergente (HCI)

- **Definición:** Integración de computación, almacenamiento y redes en una única plataforma.
- **Beneficios:** Simplificación, escalabilidad lineal, administración centralizada.

- **Soluciones:** VMware vSAN, Nutanix, Microsoft Azure Stack HCI.
- **Aplicación en digitalización:** Plataforma unificada para gestión documental.

4.5 Implementación práctica en entornos ASIR

Diseño de infraestructura virtualizada para digitalización

- Dimensionamiento de recursos según volumen documental.
- Planificación de alta disponibilidad.
- Estrategias de backup y recuperación.

Monitorización y gestión

- Herramientas de monitorización: Nagios, Zabbix, VMware vRealize Operations.
- Gestión centralizada de recursos virtualizados.
- Automatización mediante scripting y API.

Seguridad en entornos virtualizados

- Microsegmentación para protección de datos sensibles.
- Cifrado de máquinas virtuales.
- Gestión de parches y vulnerabilidades.

5. INTEGRACIÓN DE TECNOLOGÍAS EN PROYECTOS DE DIGITALIZACIÓN

5.1 Arquitectura de soluciones integradas

Componentes clave

- **Sistemas de captura:** Escáneres, cámaras, dispositivos móviles.
- **Procesamiento:** OCR, reconocimiento de formularios, extracción de datos.
- **Almacenamiento:** Repositorios documentales on-premise o cloud.
- **Gestión:** Sistemas ECM (Enterprise Content Management).
- **Automatización:** RPA, flujos de trabajo, decisiones automáticas.
- **Presentación:** Interfaces web, móviles, integración con aplicaciones corporativas.

Modelos de implementación

- **On-premise:** Control total, mayor inversión inicial.
- **Cloud:** Flexibilidad, escalabilidad, modelo OpEx.
- **Híbrido:** Combinación según requisitos de seguridad y accesibilidad.

5.2 Casos de uso avanzados

Digitalización con RPA (Robotic Process Automation)

- Automatización del procesamiento post-digitalización.
- Integración con sistemas corporativos (ERP, CRM).
- Validación automática de datos y documentos.

Digitalización con Blockchain

- Sellado temporal de documentos digitalizados.
- Verificación de autenticidad e integridad.
- Trazabilidad inmutable del ciclo de vida documental.

Uso de IA en digitalización

- **Computer Vision:** Reconocimiento avanzado de elementos visuales.
- **NLP:** Comprensión del contenido textual.
- **Machine Learning:** Clasificación automática de documentos.
- **Análisis predictivo:** Anticipación de necesidades documentales.

5.3 Metodologías de implementación de proyectos

Planificación estratégica

- Análisis de necesidades y objetivos empresariales.
- Evaluación del estado actual (As-Is).
- Definición del estado objetivo (To-Be).
- Análisis de gaps y roadmap.

Implementación ágil

- Metodologías Scrum y Kanban para proyectos de digitalización.
- Entrega iterativa e incremental.
- Priorización basada en valor.
- Feedback continuo.

Gestión del cambio organizacional

- Estrategias de comunicación y formación.
- Gestión de resistencia al cambio.
- Métricas de adopción y satisfacción.
- Liderazgo y patrocinio ejecutivo.

5.4 Evaluación y optimización continua

KPIs para proyectos de digitalización

- **Eficiencia:** Tiempos de procesamiento, costes por documento.
- **Calidad:** Precisión de OCR, tasas de error.
- **Adopción:** Uso del sistema, reducción del papel.
- **ROI:** Retorno de inversión, periodo de amortización.

Técnicas de mejora continua

- Análisis de cuellos de botella.
- Optimización de procesos.
- Actualización tecnológica.
- Benchmarking interno y externo.