

¿Cuáles son los elementos de una solución de captura automática de datos?

Aunque estemos frente a una tecnología muy compleja y avanzada, la forma en la cual se aplica es muy simple e intuitiva. Básicamente consta de 4 elementos que forman parte de todo proceso de captura automática de datos:

Los Documentos

Son la fuente de donde se extrae la información. Pueden ser papeles o imágenes, provenientes de cualquier tipo de medio. Los documentos se clasifican de acuerdo al tipo de información que contienen y a la ubicación espacial de la misma en el documento, en:

- ℳ Documentos Estructurados: el tipo de información que contienen es fijo y la ubicación de esta también es fija. Entre los documentos estructurados tenemos formularios de Censos Poblacionales, Estudios de Mercado , Evaluaciones Estudiantiles y en general cualquier documento que tenga un formato pre impreso.

- ℳ Documentos Semiestructurados: el tipo de información que contienen es fijo, pero la ubicación del contenido a extraer es desconocida. El ejemplo más claro por la aplicación actual de estas tecnologías son las facturas de proveedores. Todas las facturas deben tener en general los mismos datos, pero cada factura tiene diseños totalmente diferentes.

- ℳ Documentos No Estructurados: el tipo de información que contienen es variable y su ubicación es desconocida. En esta categoría podemos mencionar documentos como: contratos, escrituras, cartas, memorandos, etc.

[Permítanos evaluar sus documentos y le propondremos la mejor solución para su proceso.](#)

El Scanner

Es la pieza de hardware que nos permite digitalizar el documento que queremos procesar. Para esto debemos destacar que no todos los scanner están diseñados para ser usados con este tipo de tecnologías, se pueden mencionar cinco características que sin lugar a dudas son indispensables a la hora de evaluar tomar decisión por el scanner ideal que nos dará la optimización en el proceso:

Velocidad: es la cantidad de páginas por minuto que el escáner puede procesar (ppm). En el caso de escáneres de doble cara, es decir, aquellos que pueden tomar imágenes de ambos lados del papel simultáneamente, se habla de la velocidad en términos de imágenes por minuto (ipm). Las velocidades normalmente están definidas para el escaneo de papel tamaño A4 o Carta, esto significa que si el proceso es con documentos de mayor tamaño, la velocidad nominal disminuye y a su vez si el proceso es con documentos mas grandes, la velocidad nominal aumenta. Seleccione el escáner de acuerdo al papel y cantidad de documentos que desea escanear en una unidad de tiempo.

Alimentador: hay dos tipos de alimentadores, el primero y más conocido es el ADF (Automatic Document Feeder) que facilita la alimentación de los documentos que queremos digitalizar en el escáner, ya que permite apilar de una sola vez cantidades de hojas que van desde 20 hasta 1.000 papeles o más. Esta es una característica para ser evaluada en todo proceso de automatización.

El segundo tipo de alimentador es el FlatBed o Cama Plana, este alimentador solo permite que el escáner se alimente con un papel a la vez, lo que solo tiene sentido en procesos de captura automática de datos si el documento que se quiere digitalizar es antiguo, lo que lo hace muy delicado, o también si se encuentra en un estado físico tal que no sea posible alimentarlo por medio del ADF.

Tamaño: al seleccionar el escáner, es importante que se pueda conocer los tamaños mínimos y máximos que puedan soportar, de esta forma nos podemos asegurar que el tamaño del papel a ser procesado sea compatible.

Driver: Para los driver lo más importante es asegurarse que el escáner seleccionado soporte driver ISIS y TWAIN.

Formato de imagen: Por omisión el estándar usado por la tecnología de captura automática de documentos es el formato TIFF, es este el formato elegido por su versatilidad en dos características fundamentales; primero, que no pierde información, es decir todos los bits generados por el escáner pueden ser almacenados por este formato, y segundo, el tamaño del archivo, gracias a los niveles de compresión que logra.

[Permítanos evaluar su proceso y le ofreceremos el escáner más apropiado para sus procesos.](#)

El Software de Automatización

Es la pieza que toma la imagen, y sobre esta aplica las tecnologías de reconocimiento tales como OCR, ICR, OMR, que se utiliza para leer la información contenida en la misma.

Puede que en algunos casos los procesos de captura automática de datos necesiten de la intervención humana. Esto puede suceder básicamente por tres factores:

- ☞ **Calidad del papel:** el papel esta sucio, manchado o estropeado, por lo que la calidad de la imagen no es confiable.
- ☞ **Calidad del diseño:** los documentos estructurados pueden ser diseñados de tal manera que la aplicación de las tecnologías de reconocimiento automático de datos sean más eficientes, maximizando su probabilidad de lectura exitosa.
- ☞ **Calidad en carga:** hay documentos a ser procesados que se completan en forma manual, esto hace que la calidad caligráfica de la persona que lo completa conlleve a un impacto que es directamente proporcional a las tecnologías usadas para leer exitosamente la información.

Estos 3 factores definen lo que se conoce como la calidad de interpretación es el indicador fundamental a la hora de evaluar el nivel de éxito en el proceso de captura de datos, lo definimos como el porcentaje (%) de información contenida en el documento que el software pudo leer sin intervención humana.

Es importante destacar que una buena aplicación de captura de datos debe contemplar con al menos las siguientes funcionalidades:

- ☞ Validación de los datos capturados contra una base de datos. (Ejemplo: nombres, CUIT, Logo, etc.)
- ☞ Validación masiva de errores (Permite ahorrar mucho tiempo y esfuerzo en procesos de corrección)
- ☞ Validación contra imagen original (Permite visualizar la foto original del documento y sobre ella la ubicación del dato que requiere ser validado)
- ☞ Manejo avanzado de colas y lotes.
- ☞ Capacidad de procesamiento en altas cantidades de documentos.

Si lo desea lo invitamos a orientarlo para hacer su proceso más eficiente.

Los Datos de Salida

Este es el último elemento y quizás el más importante, ya que es el resultado del proceso de captura de datos, es decir, es el conjunto de datos que fueron extraídos, interpretados y verificados desde los documentos fuente.

Este conjunto de datos que puede ser catalogado como el producto final, sin lugar a dudas debe estar organizado por registros que conformado en una línea, llevarán los registros que fueron extraídos del documento.

El archivo de salida puede ser un archivo plano (.txt), un archivo XML o un archivo separado por comas (.CSV).

Como aplicación inteligente que es tiene la capacidad de alimentar directamente una base de datos, como forma de entregar los datos finales. Este es otro de los valores agregados que hace a **INVOCES y FORM** más simple y seguro el proceso de alimentación de datos a los ERP que van a usar la información recibida.

Contáctenos y lo asesoraremos en la composición de los elementos de la Captura Automática de Datos.