

PONTIFICIA UNIVERSIDAD CATOLICA DEL ECUADOR



TESIS DE GRADO PREVIA A LA OBTENCIÓN DEL TÍTULO DE MAGISTER EN GERENCIA DE
TECNOLOGÍAS DE LA INFORMACIÓN

ANÁLISIS COSTO/BENEFICIO DE IMPLEMENTAR TECNOLOGÍAS DE CAPTURA ELECTRÓNICA
DE DATOS (EDC). ÁREA DE ESTUDIO: EVALUACIONES AL TALENTO HUMANO

ESTEBAN SANSUR HOLGUÍN

INGENIERO DE SISTEMAS Y COMPUTACIÓN

QUITO, ENERO DE 2012

AGRADECIMIENTO

A Jaime Moreno y Felipe Costales, por su gran predisposición en compartir información valiosa para permitirme entender su proceso de negocio de evaluaciones psicométricas, habiendo así generado una relación comercial basada en el ganar – ganar.

DEDICATORIA

A mis padres, Miguel y Ani, por sus acertados consejos
y ejemplo de vida.

A mis hermanos Miguel y María Francisca, Santiago y
María Elena por su total apoyo incondicional y cariño.

A mis adorados sobrinos, Santiago, Mateo, Miguel
Francisco, Simón y Daniela, mi inspiración y alegría.

A todos quienes conforman el equipo de INTEL18
Ecuador por su dedicación y entrega a este gran sueño
para hacer eficiente la administración documental
digital con liderazgo, productos innovadores y servicios
de vanguardia.

1. Captura Electrónica de Datos (EDC)

1.1 Introducción

Resulta ser un hecho que la captura manual de datos es un modo bastante utilizado en organizaciones para poder extraer y explotar la información contenida en documentos físicos en diferentes procesos de negocio. La captura manual de datos es la competencia principal de las soluciones electrónicas para captura de datos EDC¹ tanto en empresas privadas y públicas.

Un número cada vez más creciente de transacciones privadas y públicas son iniciadas de manera electrónica y así mismo muchas otras son aún iniciadas mediante el llenado de formas en papel. Entonces con la pregunta ¿por qué no inician todos estos procesos de manera electrónica?, saltan a la mente varias razones fuertes para que no inician de esa manera dado que pueden tratar con un gran volumen de personas como es el caso de un censo, encuestas a peatones, en puerta o evaluaciones llevadas a cabo en lugares en donde las personas que completan la información no disponen de un computador.

Pero por qué no apoyarse en dispositivos móviles como un teléfono celular inteligente sobre el cual se desarrolle una aplicación ad-hoc para la toma de datos en sitio y evitar el manejo de papel y su procesamiento posterior? La respuesta está en la misma pregunta y recae en el desarrollo de dicha aplicación que puede resultar compleja además que demanda de recursos de alto costo para el entrevistador que no están presentes en el levantamiento con papel como el mismo dispositivo móvil, red de datos y la vulnerabilidad a que dicho dispositivo se dañe, sea extraviado o robado y con ello probablemente los datos recogidos se pierdan.

Otra razón de no usar dispositivos móviles es que la toma de datos se hace de manera individualizada, candidato por candidato, lo cual lo convierte en un sistema inaplicable si pensamos en un proceso de evaluación a varias personas en un mismo salón. Este mismo escenario podría apoyarse con el uso de computadores pero volvemos al inicio en donde

¹ EDC: Electronic Data Capture.

se requeriría una inversión en equipos de cómputo ó terminales además de la aplicación para la recolección de datos.

Organizaciones de todo tipo y tamaño se ven inundadas de formas que requieren ser procesadas y mientras que muchas de estas se apoyan en procesos manuales para la entrada de datos, existen otras que ya cuentan con los beneficios de usar tecnologías de información no solamente para automatizar el procesamiento de formas sino también para obtener exactitud, aumentar la productividad y eficiencia lo que de paso se transforma en un mejor servicio al cliente a la vez que se reducen costos. La información que contienen las formas, de cualquier manera necesitarán ser capturadas y registradas a bases de datos para así arrancar procesos como la emisión de una tarjeta de crédito, el despacho de mercadería, la emisión de una licencia o el otorgamiento de un crédito.

Eliminar el papel a través de la digitalización de documentos y crear imágenes electrónicas de formas llenas es tan solo el paso inicial de otras tareas complejas, requeridas y necesarias para transferir dicha información a bases de datos con la finalidad de ejecutar transacciones de negocio.

La premisa básica de la captura electrónica de datos radica en que un sistema informático pueda ubicar y extraer los datos desde una imagen escaneada e ingrese dichos datos en un sistema electrónico para evitar que el humano lo haga. Un sistema para captura electrónica de datos se conforma por computadores, escáneres de documentos, software especializado en reconocimiento y procesamiento de imágenes y la forma como tal. Un buen diseño de la forma así como su calidad de impresión y llenado son factores que ayudan a obtener procesos de captura eficientes, una alta tasa de confiabilidad y exactitud sobre los datos extraídos y un mayor desempeño global del sistema en sí.

En casos en donde se tratan con un número reducido de documentos y el marco de tiempo para transformar los datos contenidos en el papel físico a formatos digitales no es una restricción, frecuentemente parecería ser que resulta más óptimo digitar los datos manualmente en lugar de implementar una solución para captura electrónica de datos la cual demanda de al menos un escáner, software, la implementación de la solución y entrenamiento al personal ó más fácil aún, la contratación de un proveedor de servicios de captura electrónica de datos.

1.2 Captura de Datos y Documentos

La captura de datos consiste simplemente en la extracción de los datos contenidos en un documento de negocio, sea impreso o electrónico. La captura de datos no es lo mismo al proceso de captura de documento, ya que este último es tan solo la conversión del documento a un formato electrónico o imagen también conocido como digitalización de documentos. La captura de datos implica la conversión de la información contenida en documentos de papel o en algún formato de impresión electrónica en archivos electrónicos con una estructura consistente para su administración.

Las tecnologías de captura también implican la creación de metadata² los que describen características de un documento para posibilitar su posterior localización de manera eficiente con sistemas de búsqueda.

Los documentos pueden ser capturados por alguno de los siguientes métodos:

- Escáner de documentos.
- Importación de documentos.
- Conversión de documentos electrónicos en imágenes inalterables.

Los datos pueden ser capturados por:

- Por reemplazo de documentos en papel por formas electrónicas.
- Por extracción de información desde formas en papel como imagen mediante alguna tecnología de información para captura electrónica de datos.

1.3 Captura de Documentos

El tipo de documento a ser digitalizado determina por lo general el tipo de dispositivo de digitalización ó escáner a ser usado. Documentos como fotografías, documentos impresos a color o impresos sobre papel de color o papeles especiales pueden demandar un manejo extra antes de ser digitalizados en un escáner de cama plana o con un escáner que cuente con alimentador automático de documentos ADF³ para páginas sueltas. Si se

² Metadata: Datos respecto a los datos, considerado también como datos índices.

³ ADF: Automatic Document Feeder.

tratase de documentos sensibles físicamente ya sea por su antigüedad o por estar en libros históricos o patrimoniales que no pueden ser desarmados, el dispositivo ideal es un escáner planetario.

Una vez seleccionado el tipo de escáner, es necesario hacer una nueva selección de acuerdo al tamaño del documento, conocer si se digitalizará en modo simplex o modo dúplex⁴, un estimado de la cantidad de documentos a digitalizar cada día, calidad de los documentos y si aplica, conocer si la digitalización estará dada por la estacionalidad de un tipo de negocio o proceso.

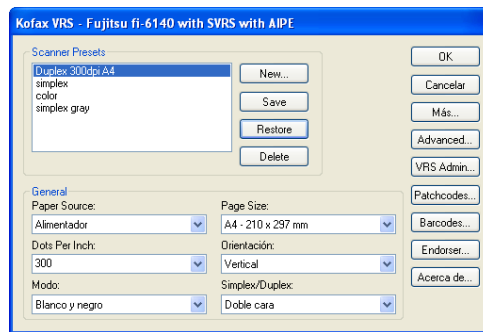


FIGURA 1-01: DIÁLOGO PARA LA CONFIGURACIÓN DE PARÁMETROS BÁSICOS DE ESCANEO.

Existen dispositivos multifuncionales MFP⁵ conocidos también como dispositivos *todo en uno* los cuales combinan escáner, copiadora e impresora con funcionalidades extendidas como conexión a una red cableada o inalámbrica y un teclado reducido ó pantalla táctil para proveer servicio de envío de correo electrónico desde el mismo equipo proveyendo a los usuarios una opción conveniente para digitalizar documentos sin necesidad de disponer un escáner conectado a la estación de trabajo del usuario o su computador. Si bien estos dispositivos multifuncionales pueden digitalizar, están diseñados para escanear documentos que se generan en procesos del día a día y no para la digitalización masiva de documentos de diversos tipos en donde se requieren de equipos robustos con volúmenes de producción diaria y velocidad de digitalización mayores a las que un dispositivo MFP puede ofrecer.

⁴ Simplex, dúplex: Capacidad de un escáner para digitalizar una o las dos caras de una página a la vez.

⁵ MFP: Multi Function Printer.

Una cámara fotográfica digital puede también ser considerada una forma de capturar documentos. Bajo ciertas condiciones óptimas de iluminación y ángulo de toma, una fotografía sobre un documento en papel puede producir una imagen con calidad suficiente inclusive para ejecutar procesos de captura electrónica de datos.

1.3.1 CAPTURA DE DOCUMENTOS POR IMPORTACIÓN

La captura de documentos por importación permite que diferentes formatos electrónicos como archivos de texto, hojas de cálculo, archivos de audio y video e inclusive imágenes o fotografías entre otros sean capturados en sus formatos nativos con el fin de ser archivados electrónicamente.

Existen soluciones que capturan y administran estos documentos electrónicos conocidos como manejadores documentales DMS⁶ o manejadores de contenido empresarial ECM⁷ Entre todas las funcionalidades con las que cuentan estos sistemas, se encuentra la opción de importar o capturar documentos electrónicos indistintos de su formato desde diferentes ubicaciones para ser almacenados en su sistema de archivos relacionándolos con un registro en una base de datos documental para ofrecer a los usuarios la posibilidad de búsqueda y acceso posterior a los mismos.

1.3.2 CAPTURA DE DOCUMENTOS POR CONVERSIÓN

De manera similar a la importación de documentos, la conversión de documentos permite que un formato electrónico modificable como por ejemplo un documento de Microsoft® Word®, una hoja de cálculo de Microsoft® Excel® ó cualquier otro formato que está siendo editado, sea convertido de manera directa en un documento electrónico no modificable como al popular formato abierto de intercambio de documentos PDF⁸. Una manera muy usual de convertir documentos electrónicos modificables a un PDF es usando una impresora genérica de PDF.

⁶ DMS: Document Management System.

⁷ ECM: Enterprise Content Management.

⁸ PDF: Portable Document Format.

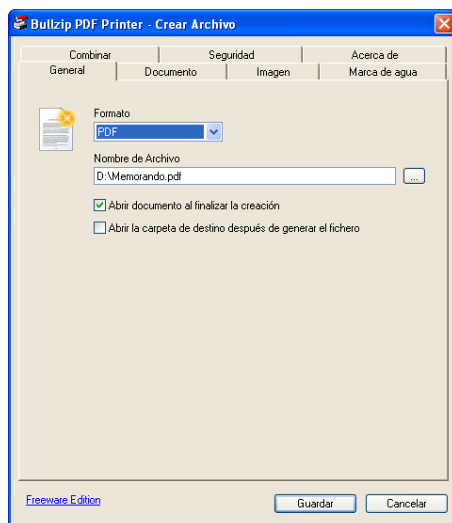


FIGURA 1-02: DIÁLOGO DE IMPRESIÓN A PDF USANDO BULLZIP PDF PRINTER

Suena contradictorio, pero el formato PDF estándar puede ser modificado posteriormente usando herramientas propias como Adobe Acrobat X Standard por mencionar tan solo una, sin embargo, PDF tiene varios sub conjuntos de especialización diseñados para cumplir estándares, entre ellos el PDF/A, diseñado para la preservación digital de documentos electrónicos garantizando que el contenido del documento es idéntico al original al momento de su reproducción.

1.3.3 CAPTURA DE DATOS DESDE FORMAS ELECTRÓNICAS

Llamada también captura dinámica de datos, permite que tanto la lógica y los datos variables con los que podrían ser completados una forma sean realizados por la simple selección de opciones por parte de quien la completa o la digitación de datos usando un teclado. Una vez completada la forma, existe un sistema asociado que interpreta las respuestas seleccionadas en cada campo de la forma para su posterior tratamiento o simplemente almacena las respuestas ingresadas en los campos.

La naturaleza de este tipo de captura requiere que la forma sea completada usando un computador sin existir otra posibilidad de entrada de información. Este requerimiento podría convertirse en algunos casos en una restricción sobre ciertos procesos masivos en donde logística, económica o geográficamente no se cuenta con acceso a computadores ó la naturaleza del levantamiento de información como tal no permiten su uso.

Un ejemplo simple de uso de este escenario es la aplicación de un crédito a una institución financiera usando una forma electrónica que el cliente accede a través del portal web de dicha institución.

1.3.4 CAPTURA DE DATOS DESDE FORMAS

Este tipo de captura, conocida también como captura estática, requiere que la forma sea impresa en papel y completada por el usuario de manera manual. Las formas en papel requieren ser digitalizadas y procesadas posteriormente con tecnologías de información para la captura electrónica de datos o en su defecto ser tabuladas de manera manual ingresando el 100% de los datos contenidos en el papel.

Un caso de aplicación de este escenario de captura son las evaluaciones de personal ó exámenes de ingreso a una institución académica, en donde el candidato llena a mano una forma que contiene las posibles respuestas.

Al ser este el tema central de la presente tesis no se amplía a mayor detalle esta introducción.

1.4 Clasificación de Soluciones según la Estrategia de Captura

Existen en el mercado diferentes soluciones para captura de documentos y datos las cuales se clasifican en tres categorías según sus funcionalidades, requerimientos de producción e infraestructura de funcionamiento. [2]

1.4.1 CAPTURA AD-HOC

Este método también llamado a demanda, es el que satisface la captura de bajos volúmenes de documentos en donde un dispositivo de captura como un escáner junto a un software especializado convierten documentos de papel o documentos electrónicos en imágenes digitales con opción de indexación y búsqueda por el contenido de cada documento. Este método puede ser orientado para trabajar con varias decenas o cientos de documentos al día y podrían incluir algún método para la administración de documentos.

Una modalidad de captura Ad-Hoc se da también con la utilización de equipos MFP, que por lo general embeben algún tipo de solución propia o de terceros en el mismo

dispositivo multifuncional para la indexación y envío del documento digitalizado a un gestor documental.

Ejemplos de soluciones Ad-Hoc de escritorio: [A]

- ABBYY FineReader Professional, Corporate.
- Canon Capture Perfect Pro.
- Nuance Omnipage.
- ReadIRIS Pro, Corporate.
- Fujitsu ScandAll Pro.
- Xerox DocuShare.

Ejemplos de soluciones Ad-Hoc sobre MFP son: [A]

- Canon UniversalSend.
- Lexmark SmartSolutions.
- Kofax Front Office Server.
- Nuance eCopy ShareScan.
- Xerox SmartSend.

1.4.2 CAPTURA POR LOTES

La captura por lotes documentales típicamente está orientada a capturar documentos desde dispositivos de escaneo de mediano y alto volumen ó ubicaciones de red, para indexarlos de acuerdo a criterios determinados y liberarlos como imágenes indexadas a algún manejador de documentos electrónicos para su publicación y facilidad de búsqueda y acceso.

El volumen de documentos que pueden capturar estas soluciones van desde cientos a miles de documentos por día, ofreciendo procesos de indexación manuales ó totalmente automatizados según la naturaleza, estructura y tipo de documento, apoyándose de, y sin limitarse a las siguientes tecnologías:

- Tecnología de reconocimiento BCR⁹.

⁹ BCR: Bar Code Recognition.

- Tecnología de reconocimiento OMR¹⁰.
- Tecnología de reconocimiento OCR¹¹ y/ó ICR¹² sobre campos fijos.
- Tecnología de reconocimiento OCR zonal asistido por el usuario (*Rubber Band OCR*).
- Tecnología de reconocimiento OCR para identificación automática de zonas basada en reglas y palabras claves.
- Tecnología de reconocimiento OCR para identificación automática de zonas por auto aprendizaje basado en documentos de ejemplo.
- Digitación manual de índices usando KFI¹³.

Adicionalmente a liberar imágenes indexadas, estas soluciones también podrían estar en capacidad de entregar los datos capturados a repositorios o bases de datos externas para apoyar procesos de negocio, convirtiéndose así en soluciones de captura transaccional la cual se explica más adelante.

La captura por lotes divide a las soluciones en dos grupos:

Soluciones de captura en red para lotes documentales [A]

Contemplada en escenarios en donde la generación de documentos así como su captura se ejecutan en puntos remotos con escáneres de mediana ó alta producción, pudiendo realizar el procesamiento de tales documentos en cada sitio remoto ó de modo centralizado donde todos los documentos capturados convergen usando alguna forma de transmisión de datos. Ejemplos de soluciones de este tipo son:

- ABBYY FlexiCapture.
- Cardiff TeleForm Enterprise.
- EMC Captiva InputAccel.

¹⁰ OMR: Optical Mark Recognition.

¹¹ OCR: Optical Character Recognition.

¹² ICR: Intelligent Character Recognition.

¹³ KFI: Key from Image.

- IRIS CapturePro
- Kofax Capture.
- Oracle Document Capture.

Soluciones de captura de escritorio para lotes documentales [A]

Usada en escenarios en donde los lotes documentales y su procesamiento es realizado en un mismo sitio usando un escáner de nivel de entrada para producción baja, media e inclusive alta. Estas soluciones podrían ofrecer métodos para indexación KFI con validación respecto al tipo de dato ingresado, validación a base de datos, indexación semi automática a partir de *rubber band OCR* y automática por reconocimiento de código de barras. Ejemplo de soluciones de este tipo son:

- Cardiff TeleForm Desktop.
- Office Gemini Diamond Vision.
- Kofax Express.
- Kodak Capture Pro.

1.4.3 CAPTURA TRANSACCIONAL

Usada para extraer datos hacia procesos de negocio con software que automáticamente identifica el tipo de documento, extrae, valida y depura los datos de interés para cada tipo documental en volúmenes que pueden comprender varios miles de imágenes al día.

Estas soluciones permiten que los procesos de extracción, validación y depuración de datos se apoyen con tecnologías OCR, ICR, BCR y OMR y puedan además usar bases de datos de negocio para así reducir la intervención manual en tareas de validación a través de búsquedas exactas o aproximadas de campos reconocidos electrónicamente para poblar otros campos con datos encontrados en dichas bases de datos. Típicamente estas soluciones se orientan para el procesamiento de formas, procesos de reclamos médicos y seguros, procesamiento de facturas y cuentas por pagar así como procesamiento de órdenes de compra, entre otros procesos más. Ejemplos de soluciones en esta categoría son:

- ABBYY FlexiCapture.
- Brainware Distiller.

- Cardiff LiquidOffice.
- IBM FileNet Capture Professional.
- IRIS CapturePro
- Kofax Capture + Kofax Transformation Modules.
- Readsoft Forms and Invoices.

Dada que la captura electrónica de datos para apoyar procesos de evaluación en talento humano es el objetivo que persigue esta tesis, la selección de soluciones y su análisis se enfoca en aquellas que permiten la captura transaccional, es decir, que puedan entregar imágenes indexadas y exportar la información capturada a bases de datos para apoyar procesos de negocio mediante el uso de las diferentes tecnologías de captura expuestas.

1.5 Indicadores y Mercados

El impulsador para digitalizar y capturar es la accesibilidad y conocimiento que obtiene el usuario al buscar y encontrar de manera ágil documentos de negocio, además de reducir costos operativos y mejorar el servicio al cliente, la captura electrónica de datos junto con soluciones para administrar documentos electrónicos introducen notables mejoras en la productividad del negocio. [1]

Según una investigación conducida por la Asociación para el Manejo de Información e Imágenes AIIM¹⁴ en Noviembre de 2010 sobre una muestra de 418 organizaciones encuestadas en los Estados Unidos, el 66% tienen un proceso formal de digitalización y archivo documental, el 47% involucra alguna forma de flujo electrónico documental en sus procesos y solo el 16% digitaliza y extrae datos hacia un proceso de negocio, es decir, que la minoría de organizaciones utilizan tecnologías de información para captura electrónica de datos y un número considerable mayor a la mitad, procesos de digitalización para solamente archivar documentos electrónicos. Tan solo el 14% de las organizaciones usa captura de datos y administración de procesos de negocio BPM¹⁵ en múltiples procesos y departamentos.

¹⁴ AIIM: Association for Information and Image Management.

¹⁵ BPM: Business Process Management.

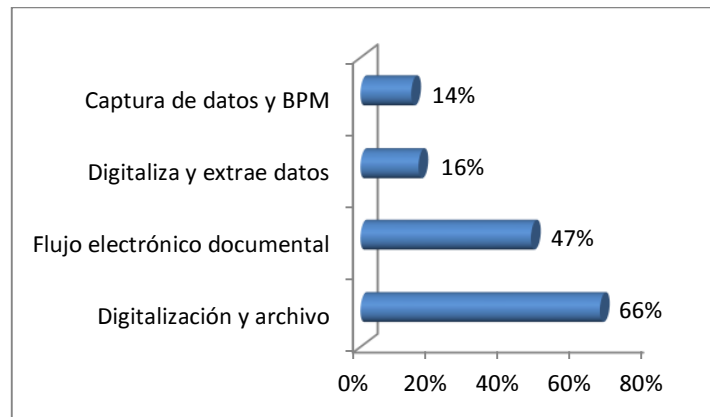


FIGURA 1-03: ORIENTACIÓN DE LA CAPTURA DE DOCUMENTOS.

La conciencia ambiental juega un papel importante dado el alto consumo de papel acompañado de la demanda y descarte de consumibles usados de impresión como cartuchos de tintas y carretes de tóner además del consumo de recursos para mantener a punto y administrar una infraestructura de impresión y copiado dentro de la organización y el manejo de estos dispositivos como basura tecnológica al cumplir su período de vida útil. Según el estudio, a pesar de que la demanda de impresión y fotocopiado creció en un 27% en el año, se detectó que en un 39% de las organizaciones, dicha demanda es menor. Este crecimiento podría estar relacionado a la expansión del mismo negocio y no necesariamente a que se requieran impresiones o copiados adicionales, pues entre las organizaciones con sistemas de escaneo y captura maduros declaran el 53% que si existe una reducción en el uso de papel apoyando iniciativas verdes.

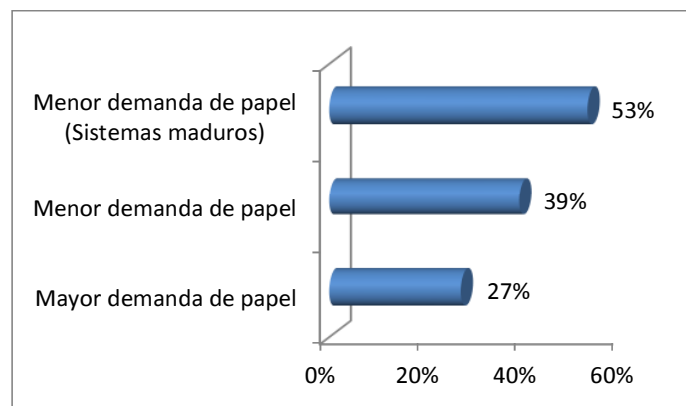


FIGURA 1-04: USO DE PAPEL.

El 55% de las organizaciones grandes (*sobre 5,000 empleados*), el 32% de mediano tamaño (*entre 500 y 5,000 empleados*) y 22% de las pequeñas (*menos de 500 empleados*) tercerizan o subcontratan servicios asociados en busca de reducir costos como razón principal (52%) y permitir que el personal se enfoque en tareas propias del negocio como razón secundaria (40%). De estas organizaciones que subcontratan una mayoría del 88% están satisfechos con los servicios que prestan sus proveedores de digitalización.

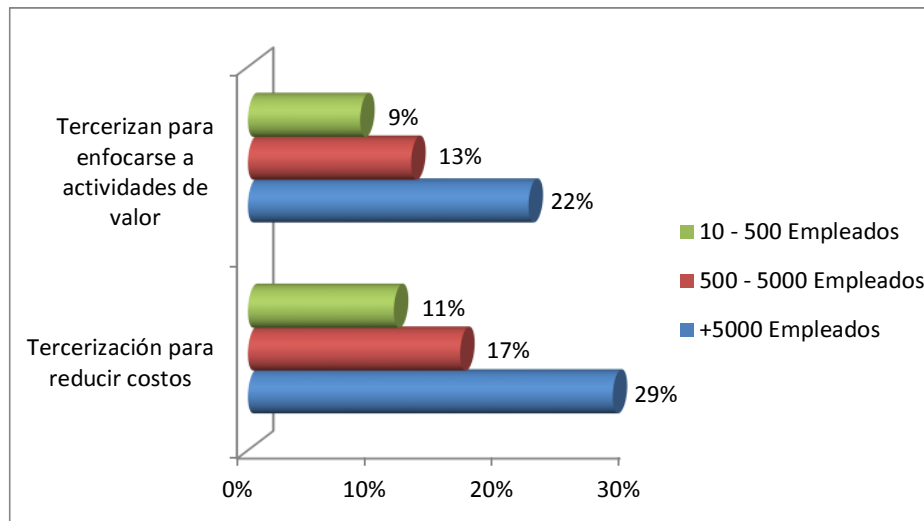


FIGURA 1-05: RAZONES PRINCIPALES PARA TERCERIZAR DIGITALIZACIÓN.

En cuanto al uso de equipos para digitalización, un 56% declaran tener escáneres de escritorio (*25% escáneres de alimentación manual y 31% escáneres con ADF*), 58% utilizan algún dispositivo multifuncional para digitalizar como parte de su proceso de archivo en el día a día, un 40% utilizan escáneres de baja y mediana producción, un 31% escáneres de alta producción encontrando que un 7% aún utilizan escáneres para microfilm.

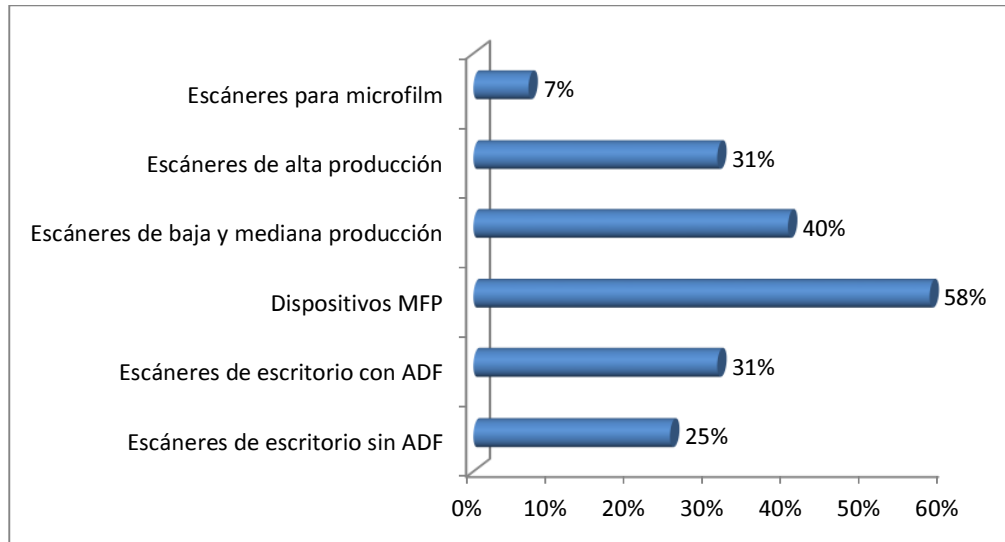


FIGURA 1-06: TIPO DE ESCÁNER UTILIZADO PARA LA DIGITALIZACIÓN.

1.5.1 ESCANEAR PARA ARCHIVAR CONTRA CAPTURAR PARA PROCESAR

A pesar de que el 82% de las organizaciones abordan la digitalización, se encontró que el 16% capturan datos para ser usados en procesos de negocio pudiendo estos documentos entrar en flujo electrónico documental pero los datos de dichos documentos fueron digitados manualmente para poder formar parte fundamental del proceso. Del 66% de organizaciones que digitalizan y archivan electrónicamente, la mitad (33%) no usan sistemas de captura de datos para automatizar la indexación y aplican metadata de manera manual para archivar electrónicamente, un 17% aplica captura de texto completo para contar con funcionalidad de búsqueda basada en contenido documental, un 14% escanean y capturan texto para automatizar el flujo documental o la auto indexación del archivo.

En cuanto a la madurez de los sistemas de digitalización y captura de datos en las organizaciones entrevistadas, un 15% usa escáneres para tareas de oficina que no incluyen integración a procesos ni archivo, tan solo un 6% digitaliza formas estructuradas y un 10% digitaliza formas semi estructuradas para capturar y extraer datos integrados a procesos de negocio. Refiérase a “Tipos de Formas” en la página 34 para mayor información respecto a formas estructuradas y semi estructuradas.

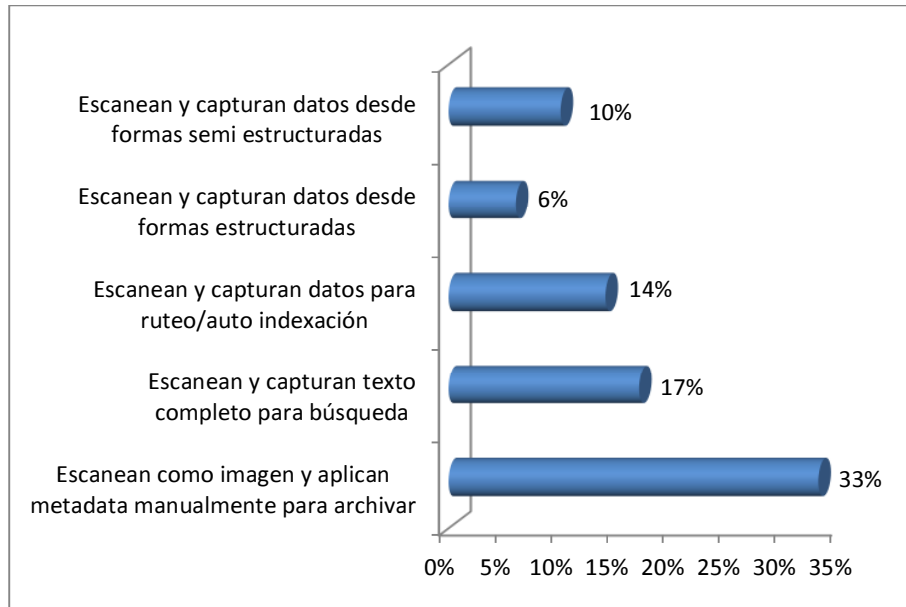


FIGURA 1-07: FINALIDAD DE ESCANEAR DOCUMENTOS Y CAPTURAR DATOS.

El sistema anfitrión de negocio que habilita procesos de captura en primer lugar es el financiero y órdenes de pedido, línea de negocio y talento humano seguido por servicios y reclamos. Si bien hay procesos de captura de datos habilitados, en los procesos financieros y de procesamiento de órdenes de pedido, un 29% integra la captura a procesos mientras que un 11% no lo integra; en talento humano se encuentra de manera pareja con captura integrada y no integrada a procesos de negocio en un 17%, y en procesos de servicios y reclamos se muestra captura integrada en un 18% y no integrada en un 10%.

Entre los principales problemas encontrados a la hora de implementar un proyecto de captura y procesos es la resistencia al cambio (38%) y la falta de percepción y conocimiento respecto a las posibilidades de un BPM (32%), siendo las dificultades de integración con otros sistemas de negocio (34%) y el tiempo de mapeo y aprobación de procesos los principales inconvenientes técnicos.

1.5.2 TERCERIZACIÓN

Los cambios tecnológicos actuales conducen a las organizaciones a analizar si sus procesos e inclusive algunos de sus sistemas anfitriones de negocio debieran ser

tercerizados con proveedores en la nube (*Cloud Computing*) en esquemas de alquiler de software como servicio SaaS¹⁶, de tal modo que todos sus esfuerzos y responsabilidad se concentran en el núcleo de negocio y actividades que generan valor apoyados en el hecho que las barreras geográficas están ya eliminadas gracias al internet y la posibilidad de contar mayor ancho de banda cada vez a menores costos. Esto sin duda mueve infraestructuras para digitalización en sitio hacia proveedores de procesamiento de documentos DPO¹⁷ dedicados y con mejores recursos, quienes pueden operar a escala global con un gran nivel de respuesta.

De las organizaciones que subcontratan servicios de digitalización, solo un 33% de estas ejecuta alguna forma de captura de datos como parte de su proceso de tercerización aunque con mayor incidencia en organizaciones grandes que en pequeñas y tan solo el 20% de los que subcontratan servicios de digitalización y captura acompañan con alguna parte del proceso de negocio, lo cual representa una excelente oportunidad para los DPO.

1.5.3 UN VISTAZO A LA CAPTURA MANUAL DE DATOS

En países en donde la hora hombre es muy económica como India, China o Filipinas, la contratación de personal para digitación es muy común para cumplir con tareas de captura de datos, las cuales generalmente incluyen:

- Pre procesamiento
 - Clasificación de documentos.
 - Análisis de documentos.
- Digitación o entrada manual de datos.
- Verificación.
- Validación de datos.

En este escenario manual de captura de datos, cada operador o digitador debe ser provisto de un lugar de trabajo, el cual implica la mayoría del gasto, dado que requiere de un computador conectado a la red de área local (*LAN*) y espacio físico.

¹⁶ SaaS: Software as a Service.

¹⁷ DPO: Document Process Outsourcers.

Según estudios realizados por ABBYY Software House, Rusia, la velocidad máxima de ingreso de datos que un digitador puede alcanzar usando un teclado estándar de distribución QWERTY¹⁸ oscila entre 250 a 400 caracteres por minuto, encontrando tasas de error de 3.5% en texto plano. Para formas (*formularios*) de cuestionarios y documentos con textos irregulares y muchos datos numéricos, la velocidad de ingreso de datos puede caer hasta o inclusive menos de 100 caracteres por minuto a la vez que el error se incrementa. [B]

Al tomar en cuenta que la productividad de un operador experimenta una reducción al acercarse a la finalización de su jornada laboral, cada uno de estos necesita períodos de receso para tomar un descanso. La media general de procesamiento para formas de mediano tamaño (*300 caracteres por página*) podría tranquilamente representar 100 formas por día, conociendo que en promedio podría tomar 5 minutos para ingresar una forma.

Las tareas de pre procesamiento de documentos demandan de personal para clasificación y personal para control de ingreso, quienes se aseguran de que no existan páginas perdidas cuando una forma contiene más de una página (*formas multipágina*) además de supervisar su clasificación. El número requerido de clasificadores y de controladores de ingreso depende directamente de la carga de trabajo esperado, siendo que en promedio, un operador de clasificación maneja hasta 1,000 formas al día y un controlador de ingreso hasta 300 formas al día.

Una vez que los datos han sido ingresados a un computador deberán ser revisados por personal de verificación, quienes verifican los datos ingresados por los digitadores y corrigen con suerte la mayoría de los errores que podrían haber ocurrido. El proceso finaliza con un supervisor quien se encarga de supervisar al equipo completo de captura de datos.

Si hablamos de proyectos medianos o grandes que requieren ser procesados en períodos de tiempo apretados, la respuesta lógica sería contratar recurso humano adicional para

¹⁸ QWERTY: Tipo de distribución de teclado.

así aumentar la capacidad de procesamiento, pero el problema es que se torna complicado aumentar la capacidad de producción de manera rápida debido a que el personal demanda de entrenamiento y por ende la capacidad de producción no aumentará de manera inmediata y proporcional al número de personas contratadas dada su curva de aprendizaje, sin contar los costos adicionales asociados a dicha contratación como equipamiento adicional, espacio de oficina más todos los beneficios al empleado. El costo de entrenamiento, contratación y reclutamiento de un empleado podría bordear el 40% del salario anual del empleado. Refiérase a “El Costo del Procesamiento Manual a Mayor Escala” en el capítulo 2 para una ejemplificación del costo de procesamiento manual.

1.5.4 EL FACTOR HUMANO

Algo sumamente crítico respecto al procesamiento manual es el factor humano el cual prácticamente es de muy compleja solución. El personal necesita ser motivado para mantener una alta productividad a lo largo de toda su jornada laboral, dado que la entrada manual de datos es un trabajo tedioso, lo que significa que inclusive digitadores experimentados cometen errores los mismos que tienden a incrementar al acercarse la finalización de su día de trabajo, lo que implica tiempo adicional para ser corregido también por operadores humanos cuya atención también podría ser deficiente. La digitación también genera tensión en la vista haciendo que la calidad de los datos ingresados sea muy baja ya que un operador no puede verificar los datos carácter por carácter durante horas seguidas.

Esta problemática generará bases de datos no satisfactorias para los clientes debido a su gran cantidad de errores y el tiempo que fue requerido para poder crearlas. Por esta razón, el procesamiento manual de formas no es la mejor opción para clientes que demandan capturar datos en mediano y gran volumen de manera regular.

Concluyendo, el procesamiento manual para captura de datos resulta ser costoso y difícilmente escalable demandando tiempo excesivo para poner en marcha al personal humano que conforma el equipo de procesamiento.

1.6 Mercados Verticales y Horizontales de Aplicación

La aplicación de tecnologías de información para la captura electrónica de datos no está diseñada exclusivamente para un tipo de negocio específico, por el contrario, puede ingresar en una gran mayoría de segmentos en donde se reciben y se generan documentos, algo que sucede en casi todo negocio.

Siendo el tema central el análisis costo/beneficio para el procesamiento de formas como servicio aplicado al talento humano, el mismo proceso podría ser replicado a otros mercados verticales por la simple razón de tratarse de formas estructuradas. Refiérase a “Tipos de Formas” en la página 34 para mayor información respecto a formas estructuradas.

En mercados horizontales el proceso puede ser ajustado de un tipo de negocio a otro satisfaciendo en diferentes segmentos los requerimientos y procesos departamentales comunes como contabilidad, ventas, compras, crédito, legal, importaciones, exportaciones y talento humano. En este caso la captura electrónica de datos podría utilizar no solo formas estructuradas sino además semi estructuradas.

Algunos mercados verticales identificados son listados a continuación y cada uno con unos pocos ejemplos de tipos de formas que pueden ser aplicados a procesos de captura electrónica de datos.

1.6.1 GOBIERNO

El sector de gobierno en general podría ser uno de los mayores consumidores de servicios BPO¹⁹ para captura electrónica de datos, dado que deben procesar una amplia variedad de formularios en grandes volúmenes tales como:

- Formas de censo.
- Encuestas en general.
- Formas de actualización e impuestos.
- Formas de registro migratorio.

¹⁹ BPO: Business Process Outsourcing.

1.6.2 LOGÍSTICA Y DISTRIBUCIÓN

Las empresas que proveen servicios logísticos y productos de distribución masiva conocen que la tarea de realizar seguimiento de documentos para cada cliente es crítico para su negocio determinando con esto capacidad de servicio y respuesta. Documentos generalmente encontrados en este tipo de negocio son:

- Órdenes de distribución y entrega.
- Solicitudes de envío.
- Manifiestos y guías de entrega.

1.6.3 EDUCACIÓN

Generalmente dos veces por año las organizaciones educativas de nivel superior reciben solicitudes de admisión de miles de aspirantes las cuales requieren ser procesadas por el departamento de admisiones. En estas organizaciones así como en las de educación básica, las pruebas de ingreso y evaluaciones corrientes demandan de gran esfuerzo para su procesamiento y calificación. Entre los tipos de formas que este sector requiere procesar se pueden mencionar:

- Formas para admisión ó inscripción.
- Pruebas de ingreso y aptitud.
- Evaluaciones periódicas aplicadas a docentes y alumnos.
- Formas para investigación y desarrollo.

1.6.4 BANCA Y FINANZAS

Cuando los clientes desean obtener tarjetas de crédito o realizar la apertura de cuentas u otros servicios, proveen muchos documentos a la institución financiera. Estos documentos así como formas diseñadas para fines específicos pueden tomar ventaja al usar captura electrónica de datos en:

- Formas para solicitud de crédito.
- Formas para solicitud de tarjetas de crédito, cuentas en general u otros servicios.
- Procesamiento de cheques.
- Identificación de documentos en general.

1.6.5 LEGAL

Tal vez es uno de los sectores con menos uniformidad en cuanto a formatos documentales por su misma naturaleza de negocio, sin embargo, la captura electrónica de datos puede generar enormes ventajas para acelerar procesos de archivo a la hora de capturar información de documentos como:

- Contratos.
- Escrituras.
- Minutas.
- Identificación de documentos en general.

1.6.6 SEGUROS

Uno de los sectores que genera gran cantidad de documentos sin duda tanto en la creación de productos y en la ejecución de procesos de siniestros. Al tratarse de documentos que cada organización entrega a sus clientes para ser completado, la agilidad con que dichos documentos sean procesados puede ser el diferenciador ante la competencia.

- Aplicaciones de seguros en general.
- Formas para declaración de siniestros.
- Identificación de documentos en general.

1.6.7 SALUD

Un sector crítico en donde la admisión de pacientes no debe esperar, la captura electrónica de datos puede apoyar el ingreso de datos de manera ágil permitiendo que sistemas de información hospitalarios HIS²⁰ puedan tener datos del paciente de manera rápida además que los documentos de ingreso pueden ser indexados y publicados a manejadores de contenido empresarial ECM ó sistemas para manejo de documentos DMS permitiendo el acceso por personal de servicio de forma precisa y oportuna.

- Registros de admisión.
- Pruebas de laboratorio.

²⁰ HIS: Hospital Information System.

- Identificación de documentos en general.

1.7 Reconocimiento de Imágenes

La digitalización de un documento, proceso con el cual, usando un escáner o inclusive una cámara fotográfica digital de alta resolución se convierte un documento en papel hacia algún formato de imagen con la finalidad de reducir la entrada manual de datos se asocia con tecnología de captura electrónica de datos.

Un factor predominante de éxito para un óptimo resultado en el reconocimiento de imágenes consiste en la selección óptima de ciertos parámetros para la digitalización del documento físico.

Un buen proceso de captura electrónica de datos será aquel que la mayoría de los datos extraídos de manera automática desde una imagen contenga la menor cantidad de errores respecto al contenido del documento original o la imagen en sí ya que esta es una representación estática del original.

Además de fijar una resolución de imagen de 300 DPI²¹ la cual es suficiente para la mayoría de documentos, existe también la profundidad de bit que también debe ser considerada con la finalidad de maximizar la eficiencia del proceso de captura electrónica de datos.

Cuando se habla de procesamiento de información, los datos que se usan como entrada determinarán en gran medida la calidad de los datos de salida, es decir, si se alimenta con datos basura se obtendrá como resultado basura. En el procesamiento de imágenes sucede lo mismo ya que el proceso de reconocimiento y el de digitalización están íntimamente ligados, por tanto gran parte del éxito depende del cuidado que se ponga al proceso de digitalización del documento.

La configuración correcta de los parámetros de digitalización permitirá que el reconocimiento de imágenes, usando alguno de los métodos que más adelante se tratan, tales como OCR, ICR, OMR u BCR gocen de mayor confianza. Estos parámetros por lo

²¹ DPI: Dots Per Inch, puntos por pulgada.

general no siempre están asociados a los que el fabricante del escáner recomienda ser los óptimos para el reconocimiento de imágenes.

1.7.1 ESCÁNER

Parte fundamental en el proceso de captura electrónica de datos. La selección de un escáner de producción que posibilite pre procesar y corregir la imagen es una característica que podría encarecer al producto pero posibilitaría ofrecer muchos beneficios al momento de la captura electrónica de datos.

1.7.2 DISEÑO DEL FORMULARIO

Factores simples que no sean considerados durante el diseño de un formulario podrían impactar notable y negativamente el desempeño de la captura electrónica de datos. La forma debería diseñarse en términos del tipo de información que se desea recolectar y la persona que la va a llenar debería contar con instrucciones claras de cómo completar la forma.

1.7.3 PROCESAMIENTO DE IMAGEN

Una funcionalidad estándar que permite aplicar pre procesamiento de imagen a nivel de software para obtener una imagen más limpia y de mejor calidad para su posterior procesamiento. Este paso es importante realizarlo al no contar con procesamiento de imagen a nivel de hardware.

1.7.4 RECONOCIMIENTO CONTEXTUAL

Permite que el contenido de un campo pueda ser reconocido ayudándose en la interpretación contextual del tipo de dato que contiene el campo, como una fecha por ejemplo, aumentando así la exactitud y desempeño del sistema de captura electrónica de datos al usar un motor de reconocimiento específico para dicho dato.

1.7.5 RESOLUCIÓN DE ESCANEADO

Al tratar de la resolución de una imagen no debe hacerse la simple relación de que mientras más alta la resolución es mejor. Antes de seleccionar una resolución, es necesario evaluar algunos puntos como:

- De qué manera serán guardadas las imágenes posterior al reconocimiento?
- La imagen será reutilizada más adelante?
- El espacio de almacenamiento es un factor a considerar?

Una respuesta rápida y efectiva es usar una resolución de 300 DPI la cual es una resolución recomendada para ejecutar procesos de reconocimiento de imágenes y datos a la vez que:

- Permite reutilizar la imagen.
- Produce un tamaño de archivo muy aceptable.
- La calidad del reconocimiento es lo suficientemente buena comparado con una resolución mayor la cual generalmente no aporta con mejoras.

Una resolución menor de 300 DPI disminuye de manera sustancial la calidad del reconocimiento de imágenes y datos, aunque producirá un tamaño de archivo menor, pero la imagen con la que se realizó el reconocimiento no necesariamente debe ser la misma que se almacena en términos de resolución, por lo que puede usarse imágenes con una resolución óptima de 300 DPI para el reconocimiento y posteriormente a este proceso transformarla a una de menor resolución, aunque no menor a 200 DPI, para obtener un tamaño de archivo menor al generado con la imagen original.

Otra razón para usar esta resolución de 300 DPI es que la mayoría de soluciones comerciales para reconocimiento de imágenes han sido diseñadas y entrenadas para dicha resolución además de que el tiempo adicional requerido para el procesamiento de una imagen en mayor resolución no compensa en lo absoluto el incremento de exactitud (si hubiere) ó la reducción en tareas de verificación manual posterior al reconocimiento.

Si bien es cierto que varios modelos de escáneres de producción de nivel de entrada de distintas marcas castigan la velocidad de escaneo cuando trabajan a una resolución de 300 DPI o más, el proceso de reconocimiento introducirá más errores que deberán ser verificados manualmente al usar imágenes de menor resolución, por lo que dicho ahorro en tiempo de digitalización no justifica en lo absoluta durante la verificación.

1.7.6 PROFUNDIDAD DE BIT

Este parámetro no solamente afecta la calidad del reconocimiento de imagen sino además en el tamaño de archivo y la reutilización de la imagen a futuro. Una imagen en color será más apetecida de usarla a futuro que una solo en blanco y negro, sin embargo, el uso de imágenes a color o en escala de gris no impactan en los procesos de

reconocimiento aunque sí en el tratamiento previo que las soluciones modernas de captura electrónica de datos realizan a las imágenes.

Una imagen en escala de gris puede resolver ciertos inconvenientes como una baja calidad en el destramado (*despeckeling*), alineación (*deskew*), y remoción de fondo (*background removal*), mientras que una imagen en color puede ser útil para procesos de reconocimiento de imágenes que no provienen de un escáner sino de una cámara fotográfica digital además de que el análisis de imágenes funciona mejor desde el punto de vista de reconocer los elementos en un documento como texto, gráficos, párrafos, líneas entre otros.

Sin embargo, cuando los procesos de reconocimiento de imágenes son asociados al tratamiento de textos y marcas, básicamente OCR, ICR, OMR y BCR, la recomendación es usar imágenes en blanco y negro, teniendo en cuenta que la usabilidad futura que se puede dar a la imagen no será la misma que si se hubiere utilizado escala de gris o color.

El tipo de imagen recomendada a usar al tratarse de OCR es TIFF²² de grupo 4, dado que este es el formato con el cual trabajan los algoritmos nativos que usan las soluciones de reconocimiento de imágenes para tratamiento de texto, ya que pueden analizar la imagen por partes sin necesidad de cargar todo el archivo en el proceso, y por ende, no sería necesario realizar una conversión previa.

Finalmente, la configuración ideal entre resolución y profundidad de color, así como otros parámetros de escaneo dependerán también en gran medida de la marca y modelo de escáner a ser usado, por lo tanto, la mejor configuración será aquella que a través de varias pruebas demuestre un mejor resultado.

1.7.7 RECONOCIMIENTO ÓPTICO DE CARACTERES (OCR)

Es la tecnología que permite a una solución de captura convertir diferentes tipos de documentos en papel que han sido escaneados o formatos electrónicos no editables en datos posibles de ser interpretados por un computador de tal manera que el usuario puede usar dichos datos o texto digital para edición o búsqueda posterior.

²² TIFF: Tagged Image File Format.

El reconocimiento óptico de caracteres básicamente es una tecnología de correspondencia de patrones, la cual interpreta la figura y patrón de caracteres que se encuentran impresos sobre un documento generado por una impresora, a su equivalente código de computador, lo que a su vez representa un carácter digital como letras, números o símbolos que tanto las computadoras y el humano pueden usarlo.

Los humanos reconocemos un carácter por la variación en su figura y patrón, inclusive cuando el mismo carácter tiene diferentes patrones lo asociamos con uno solo, tal como se ilustra a continuación con el carácter “A”:

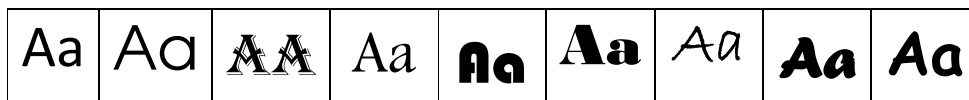


FIGURA 1-08: VARIACIONES MAYÚSCULA/MINÚSCULA DE LA LETRA “A” EN DISTINTAS TIPOGRAFÍAS.

Un computador para poder interpretar digitalmente cada carácter lo interpreta por su código binario asociado, traduciendo y presentando al humano como el carácter apropiado estándar que es interpretable para este. Adicionalmente, al existir diferentes tipografías, el cerebro humano es capaz de reconocer, interpretar y comprender todas las variaciones de manera extremadamente rápida. De una manera análoga, el computador ejecuta una serie de algoritmos muy intensivos para intentar simular lo que el cerebro humano realiza en fracciones de segundos, traduciendo un carácter presentado como imagen, conocido como mapa de bits, en su correspondiente carácter entendible para el computador y el humano.

1.7.8 RECONOCIMIENTO INTELIGENTE DE CARACTERES (ICR)

Es la tecnología que permite convertir caracteres escritos a mano dentro de campos y marcos a caracteres posibles de ser interpretados por un computador.

La tecnología ICR es más compleja de ejecutar que el OCR dado que:

- No todas las personas escriben de la misma manera.
- Una misma persona puede escribir un mismo carácter de diferente manera en diferentes situaciones y circunstancias.
- Los caracteres no están siempre espaciados equitativamente ó inclusive pueden estar enlazados.

La tecnología ICR es la que se aplica en su mayoría al tratar con el procesamiento de formas y requiere que la escritura manual empleada durante el llenado use caracteres de molde o letra tipo imprenta para lograr los beneficios para lo que fue desarrollado, cumpliendo las siguientes consideraciones:

Texto con espaciado uniforme

El espacio que existe entre cada uno de los caracteres dentro de un campo debe mantener la misma relación uno de otro. Esto se logra creando marcas o cajones contiguos dentro de los cuales se escribirá un carácter.

Uniformidad en ancho y alto

Cada carácter en lo posible trata de tener el mismo alto y ancho, de esta manera el proceso de ICR será más preciso al reconocer cada uno de los caracteres.

Línea base uniforme

El texto escrito en letra de molde debe estar sobre la misma línea horizontal de base causando una correcta segmentación y reconocimiento de caracteres con colas como las letras “p”, “q”, “b” y “d”.

Guía de llenado

Un mapa de caracteres modelo incluido en la forma a ser completada será de gran utilidad para mostrar cómo se representan los caracteres que el ICR reconocerá con una alta confiabilidad.

1.7.9 RECONOCIMIENTO ÓPTICO DE MARCAS (OMR)

Es la tecnología que detecta la presencia o ausencia de una marca en un área determinada. El reconocimiento óptico de marcas es usado con gran frecuencia en formas para recolectar información a partir de preguntas de selección única o múltiple en donde la respuesta se puede dar por simple elección.

Según el tipo de sistema de captura electrónica de datos, la forma de dicha marca puede ser relevante o no, es decir, aceptar marcas de cualquier forma como X ó ✓, ó solo del tipo burbuja en la que se debe pintar totalmente el área destinada a la marca para denotar su selección.

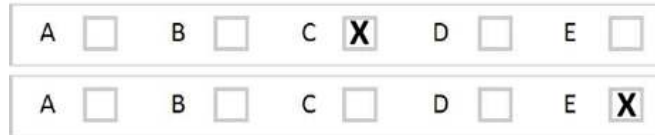


FIGURA 1-09: MARCAS DE SELECCIÓN TIPO "X".

El reconocimiento óptico de marcas OMR puede funcionar con dos tipos de tecnología para su lectura: La primera usando un lector óptico de marcas para interpretar formas OMR y la segunda utilizando un escáner de imágenes apoyado de software para lectura de OMR.

1.7.10 OMR POR MEDIO DE LECTOR ÓPTICO DE MARCAS

Es un tipo de escáner que se encarga de capturar e interpretar la presencia o ausencia de marcas en áreas predeterminadas sobre una forma OMR. El área de estas marcas conocidas como burbujas, deben ser pintadas totalmente para denotar la selección realizada por la persona que llena la forma OMR. La siguiente figura presenta la selección del número 09511 que ha sido realizada sobre una forma OMR.

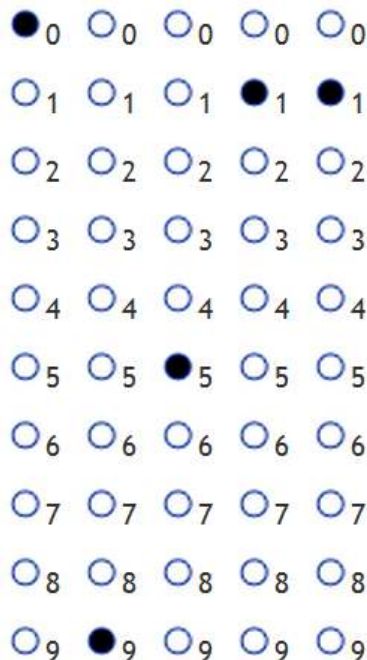


FIGURA 1-10: MARCAS DEL TIPO BURBUJA SOBRE UNA FORMA OMR.

Las características de un sistema de lector óptico de marcas son:

Velocidad

Dependerá de la velocidad a la que el lector pueda escanear cada página pudiendo encontrarse en el mercado equipos que leen sobre 60 páginas por minuto ó más. En sí la tarea de procesamiento es realizada por el mismo lector por lo que prácticamente pueden trabajar en modo desatendido sin intervención humana cuando la forma OMR ha sido completada de manera correcta.

Exactitud

Un escáner OMR mide el nivel de oscuridad encontrado en una burbuja para poder determinar si un espacio de selección ha sido o no marcado, ofreciendo niveles de exactitud del 99.9%

Costo

Una solución OMR basada en lector óptico es generalmente más costosa que una solución basada en escáner de imágenes, además que el lector en sí es destinado únicamente para dicho fin.

Limitación

Un lector óptico no puede reconocer o procesar datos alfanuméricos mediante tecnología OCR e ICR por lo que solo se limita al reconocimiento de marcas.

Método de llenado

Una burbuja para ser reconocida como seleccionada debe estar pintada prácticamente en su totalidad. Generalmente estas formas OMR para captura por lector óptico incluyen una sección con ejemplos de llenado correcto e incorrecto, de modo similar al mostrado en la figura siguiente:



FIGURA 1-11: EJEMPLO DE INSTRUCCIONES DE MARCADO EN UNA FORMA OMR PARA LECTOR ÓPTICO.

Otras consideraciones

- Una forma OMR para lector óptico puede contener burbujas unas muy cerca de otras, haciendo posible la recolección de datos en un menor espacio de página.
- Una forma OMR requiere ser diseñada con extremo cuidado y exactitud.
- La forma OMR no puede ser foto copiada, debe ser original.
- Los lectores ópticos no pueden manejar tolerancias de inexactitud, por lo cual los formularios deben ser 100% precisos respecto a la posición relativa de las marcas a ser leídas, es decir, solamente utiliza formas estructuradas o fijas.
- Dependiendo del modelo de lector óptico puede manejar una bandeja independiente para documentos que son rechazados.
- El lector óptico no puede ser usado para escanear documentos que no sean formas exclusivamente diseñadas para OMR de lectura óptica.

1.7.11 OMR POR MEDIO DE ESCÁNER DE IMÁGENES

Es un método que usa un escáner común de escritorio o dedicado de volumen con ADF. El escáner de imágenes como tal no es quien se encarga del procesamiento de las formas digitalizadas, sino que estas precisan de software especializado para la captura y procesamiento de marcas a partir de las imágenes entregadas por el escáner.

Los formularios diseñados para ser procesados con este tipo de escáner y software pueden ser creados de manera ágil y rápida brindando flexibilidad de impresión y un alcance más amplio en procesamiento que no se limita solamente a formas OMR.

Las características de un sistema para captura de marcas usando escáner de imágenes son:

Velocidad

La velocidad de digitalización de un escáner de imágenes varía según el modelo y marca, sin embargo, escáneres de escritorio de hasta US\$ 600 dólares americanos pueden rendir 20 páginas por minuto en modo simplex, ó 60 páginas por minuto con escáneres que bordean los US\$ 2,000 dólares americanos. Cabe recalcar que existen escáneres de producción que alcanzan 100 páginas por minuto en modo simplex (200 imágenes por minuto en modo dúplex) con ciclos de producción diaria

de 100,000 páginas, cuyo costo puede ser similar o superior a un lector óptico de marcas. [A]

Exactitud

A pesar que los escáneres de imágenes podrían proveer una exactitud levemente inferior que un lector óptico, existen factores que logran dramáticas mejoras en el procesamiento de imágenes y captura electrónica de datos así como el diseño en sí de la forma usada para capturar los datos. La mezcla inteligente de estos componentes puede lograr como resultado una exactitud del 99.9% al tratar con formas OMR.

Adicionalmente un escáner de imágenes es más sensible por lo que puede leer tonos de gris y no solamente discriminar entre blanco y negro como un lector óptico de marcas.

Costo

Un escáner de imágenes de escritorio es un equipo más económico que un lector óptico de marcas. A este costo de escáner se adicionará el costo del software para el procesamiento de formas OMR.

Funcionalidad del escáner

El escáner puede ser utilizado para la digitalización de otros documentos y no solamente para la digitalización de las formas OMR, con lo cual, un mismo equipo es usado para abarcar la digitalización en general.

Otras consideraciones

- Una forma tipo OMR podría requerir de mayor espacio en blanco entre las marcas destinadas para la selección, aunque actualmente se pueden encontrar soluciones modernas de captura electrónica de datos que logran eliminar esta necesidad de mayor espacio entre las marcas, pudiendo igualar a las que se presentan en una forma OMR para lector óptico.
- Un campo para ser reconocido como seleccionada puede estar pintado completamente ó contener marcas del tipo X, ✓, +, / y otras.

- Una forma a ser procesada con un escáner de imágenes podría ser foto copiada.
- Los formularios para escáner de imágenes no demandan de gran exactitud respecto a la ubicación absoluta de los campos, ya que el software puede corregir distorsiones durante el procesamiento de imagen previo al reconocimiento.
- Si el software lo permite, una forma para escáner de imágenes puede ser usada no solo para capturar marcas OMR, sino también datos alfanuméricos impresos por computador ó llenados a mano usando letra de molde. El software especializado procesará estos campos aplicando tecnología OCR e ICR, brindando mayor flexibilidad y alcance para diferentes proyectos o estudios de captura de datos sin limitarse a la captura de marcas. Adicionalmente, también pueden interpretar códigos de barras de una y dos dimensiones con una confiabilidad del 100%.

1.7.12 RECONOCIMIENTO DE CÓDIGO DE BARRAS (BCR)

Los códigos de barras son por naturaleza una representación de datos diseñada para ser leída por computador o por lectores especializados. Existen códigos de barras de una dimensión o lineales (*1D*) los que están conformados por líneas paralelas de diferente ancho y espacio de separación entre sí, y de dos dimensiones (*2D*) caracterizados por contener y representar una mayor densidad de datos.



FIGURA 1-12: EJEMPLO DE CÓDIGO DE BARRAS 1D TIPO EAN128 PARA EL TEXTO "01234567890".



FIGURA 1-13: EJEMPLO DE CÓDIGO DE BARRAS 2D TIPO PDF 417 (ISO 15438) PARA EL TEXTO "01234567890".



FIGURA 1-14: EJEMPLO DE CÓDIGO DE BARRAS 2D QR (QUICK RESPONSE ISO 18004) PARA EL TEXTO "0123456789".

El código de barras en procesos de captura electrónica de datos puede ser usado como identificadores de formas o marcas de referencia para agilizar y garantizar la exactitud en procesos de selección y clasificación de documentos o formas. Adicionalmente, un código de barras también puede ser usado para la creación y separación de lotes automática de documentos durante el escaneo de formas cuando se trabaja con medianos o altos volúmenes de documentos, teniendo la posibilidad de que el software se encargue de identificar y clasificar documentos según el código encontrado y procesarlos de manera independiente y diferente a cada tipo, eliminando la necesidad de que las formas deban ser clasificadas de manera manual previamente al escaneo.

1.8 Formas para Procesamiento por Computador

A diferencia de las cartas, correos electrónicos, artículos en un periódico ó un documento cualquiera con formatos libre, una forma se caracteriza por presentar un formato consistente en todo su esquema o *layout*.

En una forma existen áreas específicas para cada tipo de información que es requerida conocer, tal como un nombre o un número de cédula, los que son consistentemente ubicados.

Principalmente una forma difiere de otro tipo de documento por las siguientes dos razones:

1. Una forma siempre tiene una cantidad predeterminada y conocida de campos.
2. Cada campo puede contener un solo tipo de dato, por ejemplo: Un campo para el número de cédula solo puede contener caracteres numéricos entre el cero y el

nueve con una longitud fija de 10 caracteres, mientras que un campo de Nombre solo contiene caracteres alfabéticos con longitud variable.

El procesamiento automático de formas consiste básicamente de las siguientes etapas:

1. Diseño de un esquema de forma el cual especifica y caracteriza de manera única a una forma a la vez que define los campos a ser reconocidos.
2. Llenado de las formas y posterior digitalización.
3. Procesamiento de las imágenes, lo que incluye:
 - a) Identificación de forma.
 - b) Reconocimiento de campos.
 - c) Verificación y validación de datos.
4. Exportación de los datos extraídos a almacenes externos de información o bases de datos.

Al iniciar el procesamiento, una forma primero debe hallar correspondencia usando zonas de OCR y BCR con solo uno de los esquemas de entre todos los que el sistema de captura electrónica de datos podría usar como alternativas para la identificación de la forma. Una vez identificado el esquema correspondiente, podrá realizar la discriminación entre aquellos elementos guías que no requieren reconocimiento, tales como bordes de campos, cadenas de texto, líneas, fondos e imágenes y los campos para reconocimiento para la extracción de datos sobre los que deberá aplicar OCR, ICR y OMR.

Según el nivel de confianza reportado posterior al reconocimiento de los campos, se ejecutan procesos manuales de verificación y validación de datos para garantizar la exactitud de los mismos antes de ser exportados.

1.8.1 TIPOS DE FORMAS

Los documentos según su naturaleza se clasifican en tres categorías:

1. Formas estructuradas.
2. Formas semi estructuradas.
3. Formas no estructuradas.

Formas estructuradas

Se entiende por un documento o forma estructurada aquellos que tienen campos de datos dedicados los cuales permanecen constantes tanto en cantidad, posición absoluta y formato durante todas las copias del documento. Estas formas suelen ser emitidas en formatos impresos para posteriormente ser completadas a mano.

Una forma estructurada es diseñada para un fin específico, generalmente de investigación, con el que se recolecta información de manera masiva como encuestas, censos, cuestionarios o evaluaciones, cuya información recolectada luego de ser ingresada a una base de datos es tratada y procesada matemática y/o estadísticamente para la toma de decisiones o la simple publicación de resultados.

Una forma estructurada también es conocida como forma fija.

Formas semi estructuradas

Un documento o forma semi estructurada es aquella cuyos campos de datos difieren en cantidad, posición relativa y formato entre las diferentes instancias del documento.

Un ejemplo claro de una forma semi estructurada es una factura comercial, dado que su formato así como el número de ítems difieren de cada emisor, aunque todas las facturas tienen en común un código o identificador de cliente, un número de factura, una fecha de emisión, un monto total y otros datos adicionales que están localizados en diferentes posiciones sobre el documento. Estos datos posteriormente a ser ingresados en una base de datos arrancan o apoyan procesos de negocio, como en el caso de la factura, iniciaría un proceso de pago.

Una forma semi estructurada se la identifica también como forma flexible.

Formas no estructuradas

Se refiere a cualquier forma que no cae en ninguna de las categorías anteriores englobando a todos aquellos documentos de forma libre como contratos, cartas, diagramas u otros los que pueden ser tratados de manera individual o como soportes y/o complementos de algún otro documento fijo o flexible. Sobre las formas no estructuradas se ejecutan procesos de OCR completo y se extraen unos

pocos índices manuales ó basados en reglas para habilitar su búsqueda por atributos o texto completo desde sistemas de gestión documental.

1.9 Elementos de una Forma

Los siguientes elementos son encontrados en una forma.

1.9.1 TEXTO

Elemento de una forma el cual contiene texto descriptivo como el título de la forma, nombres de campos, explicaciones u otros. La tecnología de reconocimiento usada en elementos de texto es OCR.

Al tratarse de datos pre impresos que deban ser reconocidos, se recomienda usar un tamaño de letra de 10 puntos o más con espacios fijo entre caracteres y mayúsculas de preferencia. Si son textos explicativos que no serán reconocidos usando OCR, la selección de tamaño y tipo de letra queda a criterio del diseñador de la forma, aunque se podría optar por estandarizar para lograr una presentación visual agradable.

La tipografía recomendada para elementos de texto es Courier, Courier New, OCR A y OCR B. El juego de caracteres de las tipografías mencionadas tienen la misma relación de dimensión horizontal independientemente del carácter que representa, lo que no sucede por ejemplo en la tipografía Arial cuya relación horizontal no es la misma ó Times New Roman en la que puede ser difícil distinguir la “m” de “rn”. Refiérase al Anexo 1 para consultar el mapa de caracteres alfanumérico de las tipografías señaladas.

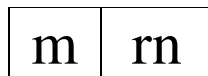


FIGURA 1-14: SIMILITUD GRÁFICA EN LA TIPOGRAFÍA TIMES NEW ROMAN ENTRE “M” Y “RN”.

La tipografía OCR-A fue desarrollada en 1968 específicamente para ser usada de manera eficiente en soluciones de OCR y es un formato estandarizado por ANSI²³. La tipografía OCR-B también desarrollada en el mismo año para el mismo fin presenta trazos de caracteres menos técnicos que OCR-A. Estas tipografías garantizan el reconocimiento de

²³ ANSI: American National Standards Institute.

texto con una exactitud que excede el 99%, dándose una tasa de sustitución de datos de una en cien mil, es decir el 0,001%. [3]

1.9.2 CAMPO PARA ENTRADA DE DATOS

Es un elemento básico dentro del cual la persona que llena la forma lo usa para escribir texto. La tecnología de reconocimiento usada en estos elementos es ICR.

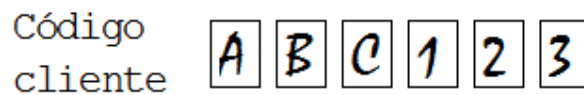


FIGURA 1-15: REPRESENTACIÓN DE UN CAMPO PARA ENTRADA DE DATOS.

Para facilitar la entrada de texto, estos campos se delimitan con distintos estilos de marcos acompañados por textos explicativos que describen los datos con los que deben completarse el campo. Los distintos estilos que pueden ser usados para la delimitación de marcos dependerá mayoritariamente de lo que sugiera el fabricante de la solución de captura, y probablemente también a temas de diseño y presentación visual.

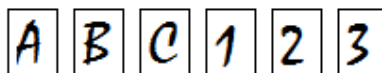


FIGURA 1-16: TEXTO EN MARCOS INDEPENDIENTES O SEPARADOS.

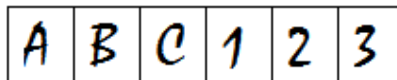


FIGURA 1-17: TEXTO EN MARCOS CONTIGUOS.

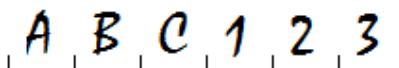


FIGURA 1-18: TEXTO SOBRE GUÍA TIPO PEINE.

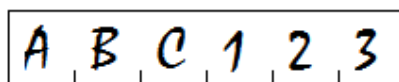


FIGURA 1-19: TEXTO SOBRE GUÍA TIPO PEINE CON MARCO.

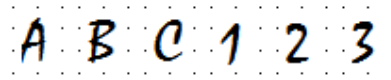


FIGURA 1-20: TEXTO EN MARCOS DE TRAMA.

1.9.3 CAMPO DE SELECCIÓN

Es un elemento frecuentemente encontrado en una forma en donde se espera algún tipo de marca manual realizada por el usuario que llena el formulario para indicar su selección. Un campo de selección asocia un texto que explica la equivalencia de selección de dicho campo. La tecnología de reconocimiento usada en campo de selección es OMR.

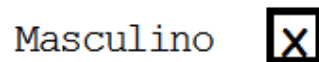


FIGURA 1-21: EJEMPLO DE CAMPO DE SELECCIÓN.

A pesar que la figura muestra un cuadrado como campo de selección, es posible diseñarlo como un círculo, un polígono, una línea baja o inclusive un espacio vacío donde se receipte la marca. El campo de selección que se deba usar no solo obedecerá a temas de diseño sino también a la preferencia del software de captura para lograr una alta confiabilidad en el reconocimiento de marcas.

1.9.4 GRUPO DE CAMPOS DE SELECCIÓN

En un elemento que agrupa dos o más campos de selección en un conjunto lógico de posibles respuestas acompañados de un texto explicativo para denotar el contexto del conjunto. La selección realizada en el grupo puede ser de respuesta única o múltiple y es el mismo contexto de pregunta o alguna indicación explícita la que instruirá el tipo de respuesta.

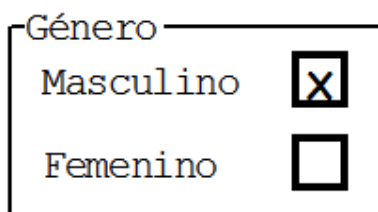


FIGURA 1-21: AGRUPACIÓN LÓGICA PARA CAMPOS DE SELECCIÓN.

1.9.5 *MARCAS DE REFERENCIA*

Conocidas también como marcas de anclaje, son un elemento fundamental en una forma estructurada las cuales ayudan en su conjunto al sistema de captura para:

- Identificar la correspondencia de la forma digitalizada con la plantilla que el sistema debe usar para el procesamiento.
- Determinar la orientación de la forma, vertical u horizontal.
- Corregir ligeras variaciones de desplazamiento de imagen causadas por la inexactitud del proceso de impresión, copiado o por la desalineación/distorsión causada por el alimentador de páginas del escáner al momento de digitalizar.
- Servir de referencia para ubicar la posición exacta de los campos a ser reconocidos.

Todas las marcas de referencia que están presentes en una forma fija deben:

- Encontrarse en la misma posición absoluta para un mismo tipo de forma.
- Formar un rectángulo lógico entre al menos 4 de estas. Una quinta marca de anclaje puede ubicarse en cualquiera de los lados de este rectángulo, e inclusive dentro sin que se interponga a algún campo a capturar.
- Ser de tamaño único y más grande que el carácter promedio de texto.
- Ser impresas en color negro y ser del mismo tipo.
- Mantener un área despejada de al menos 25mm en todos sus lados.

Los tipos de marcas de referencia que pueden usarse pueden ser cuadrados o rectángulos sólidos, cruces, esquinas e inclusive líneas o un texto específico. Ejemplos:



FIGURA 1-22: TIPOS DE MARCAS DE ANCLAJE.

1.9.6 LÍNEAS SEPARADORAS

Es un elemento en forma de una línea horizontal o vertical de color negro y son usadas para dividir una forma en varios segmentos lógicos. Estas líneas podrían ser usadas también como marcas de anclaje aunque su aplicación se relaciona más al diseño de la forma.

1.9.7 IDENTIFICADOR DE FORMA

Un identificador de forma es un elemento que es usado para seleccionar la plantilla adecuada de procesamiento en el caso de que varias plantillas tengan la misma distribución de marcas de referencia.

Es recomendable usar como identificador de forma un código de barras, un número único ó título con alguna tipografía óptima para OCR.

1.9.8 IMAGEN

Un elemento definido como una imagen es una sección a la cual no se aplica reconocimiento y podría simplemente ser usada para exportarla como un archivo de imagen, como puede ser el caso de un espacio destinado para la firma de la persona que completa la forma o un sello.

Algunos sistemas sofisticados de procesamiento de imágenes podrían aplicar algoritmos para la comparación de imágenes, como sucede en el caso de las firmas. La finalidad es que el sistema genere un veredicto de que una firma coincide con la registrada en una base de datos de firmas.

1.9.9 TABLA

Es un arreglo de columnas separados con líneas y títulos en la cabecera de cada columna las cuales contienen el mismo tipo de elementos como textos, campos para entrada de datos, fechas, números, campos de selección. Un ejemplo práctico de una tabla es la sección de detalle de una factura.

1.10 Consideraciones y Diseño

Una forma bien diseñada acelera el procesamiento e incrementa la confiabilidad durante la captura electrónica de datos, de tal modo que para alcanzar la eficiencia en la extracción de datos es mandatorio crear un esquema correcto.

De manera resumida, una forma debería cumplir con las siguientes características:

1. Ser simple de completar por parte del usuario.
2. Usar pocos métodos para la recolección de información tales como:
 - a) Preguntas de selección con opción múltiple.
 - b) Preguntas de selección con opción única.
 - c) Pocas preguntas para respuestas cerradas.
 - d) Muy pocas preguntas para respuestas abiertas.
3. Los campos de datos deben estar claramente identificados y explicados para que las respuestas tengan el formato correcto.
4. Las instrucciones deben estar claramente escritas y con lenguaje simple.

El esquema general de una forma y su diseño debe ser lo más amigable posible tanto para los usuarios como para los sistemas de reconocimiento y debería en un inicio identificar la información que se va a capturar, los campos que se usarán para entrada de datos y el esquema físico de la forma.

Sabiendo que en una forma coexisten áreas que no deben ser reconocidas con áreas que si deben reconocerse, y que estas deben ser fácilmente distinguibles entre sí, el diseño puede lograrse de dos maneras:

1. Haciendo que los bordes de campos, fondos y marcas de texto desaparezcan o se desvanezcan mientras la forma es digitalizada.
2. Evitando que los bordes de campos, fondos y marcas de texto desaparezcan mientras la forma es digitalizada.

La primera opción hará que solo las marcas de referencias, textos explicativos y los datos escritos en la forma permanezcan intactos durante la digitalización de tal modo que la solución de captura electrónica de datos no tendrá que separar los datos escritos de los elementos de la forma lo cual hace que el reconocimiento sea más fácil, rápido y confiable.

La segunda opción hará que las marcas de referencias, textos explicativos, los datos escritos así como las marcas de texto y los bordes de campos permanezcan intactos durante la digitalización. En este caso la solución de captura electrónica de datos buscará

la información que está rodeada por los bordes de campos y marcas de texto. La ausencia de estos elementos de forma causará un impacto negativo en la calidad de reconocimiento y aumentará el tiempo de procesamiento.

Las variantes de diseño para formas cuyos elementos se desvanecen durante la digitalización son:

1. Formas para segregación de color (*Color dropout forms*).
2. Formas blanco y negro con fondos de trama (*black and white forms with raster backgrounds*).
3. Formas blanco y negro con bordes de trama (*black and white forms with raster borders*).
4. Formas blanco y negro lineales (*black and white linear forms*).

1.10.1 FORMAS PARA SEGREGACIÓN DE COLOR

Una forma para segregación de color se diseña usando ciertos colores en tonalidades que son visibles para el humano pero que no lo son para un escáner al ser digitalizadas en modo blanco y negro, debido a que la reflexión de color de estos tonos es por lo general mayor al 60%. Estos colores son usualmente tonalidades claras de rojo, naranja, verde e inclusive gris. Estas formas usan un color suave de fondo con rectángulos blancos los que sirven como celdas para cada carácter, campo de selección u otro. Refiérase al Anexo 2 para obtener el listado de códigos Pantone para los colores posibles de usar en formas para segregación de color.

De manera ideal, todos los elementos en una forma a excepción de las marcas de referencia deberían desaparecer durante el escaneado, dejando solamente el contenido de los campos para ser reconocidos.

Los textos explicativos podrían ser retenidos si son impresos en negro, pero para que dichos textos desaparezcan en la digitalización pueden ser impresos con el mismo color de fondo pero con una mayor saturación de color.

A pesar que el color verde es un color que puede ser usado en una forma para segregación de color, se recomienda usar el rojo ó el naranja, ya que estos tienen mayor contraste respecto al color azul, no así el verde, por lo que se obtendrá buenos resultados

de reconocimiento si la forma fue completada con tinta azul. Refiérase al Anexo 3 para visualizar un ejemplo de forma para segregación de color.

1.10.2 FORMA BLANCO Y NEGRO CON FONDOS DE TRAMA

Los campos en este diseño de forma son espacios en blanco simples, usualmente rectángulos, sobre un fondo de trama creado con puntos individuales.

El tamaño recomendado de cada punto no debe superar a 0,1mm con una distancia entre cada punto de 1mm aproximadamente. Estos parámetros son cruciales ya que este fondo de trama será eliminado durante el proceso de destramado en lugar de eliminarse durante la digitalización. Este proceso se encarga de tratar a dichos puntos como basura o ruido.

1.10.3 FORMA BLANCO Y NEGRO CON BORDES DE TRAMA

Estas formas carecen de fondo y todos los bordes de campos, campos de selección y otros elementos son hechos a partir de tramas o una secuencia de puntos negros pequeños cuyo tamaño recomendado es de 0.39 puntos con una distancia entre punto y punto de al menos cinco veces el tamaño del punto. La remoción de estos puntos tiene lugar al aplicar el filtro de destramado. Refiérase al Anexo 4 para visualizar un ejemplo de forma blanco y negro con bordes de trama.

1.10.4 FORMA BLANCO Y NEGRO LINEALES

Estas formas carecen de fondo en la mayoría de casos. Todos los bordes tanto para campos de texto, campos de selección y otros elementos están hechos de líneas solidas en color negro. Estas formas deben llenarse con nitidez caso contrario las letras podrían traslapar los bordes dificultando la separación de los bordes y el texto reduciendo así la calidad de reconocimiento. De manera ideal, esta forma debería completarse tratando en lo posible de que la escritura no tope a ningún borde. Refiérase al Anexo 5 para visualizar un ejemplo de forma blanco y negro lineal.

1.10.5 SELECCIÓN DEL DISEÑO DE FORMA

Junto con los requerimientos de procesamiento para cada diseño de forma, es necesario tener en cuenta el costo y método de impresión de cada variante, el número de copias requeridas, el tamaño resultante de imagen digitalizada, el modo de escaneo, el nivel requerido de exactitud de reconocimiento, la velocidad de procesamiento, la apariencia

ANÁLISIS COSTO/BENEFICIO DE IMPLEMENTAR TECNOLOGÍAS DE CAPTURA ELECTRÓNICA DE DATOS (EDC). ÁREA DE ESTUDIO: EVALUACIONES AL TALENTO HUMANO

en general de la forma y el cuidado que se requiere durante el llenado de la forma. A continuación se resumen las ventajas y desventajas respecto a factores críticos en cada variante de diseño:

Criterio de Evaluación	Variante de Diseño de Forma					
	Segregación de color		Trama blanco y negro		Blanco y negro lineal	
	Desventaja	Ventaja	Desventaja	Ventaja	Desventaja	Ventaja
Complejidad en Diseño		Fácil de diseñar en cualquier editor gráfico	Un editor de texto tiene limitaciones para diseñar este tipo. El tamaño de los puntos de trama es crucial.			Fácil de diseñar en cualquier editor gráfico
Impresión	Complejo de imprimir en alta calidad con equipo de casa			Fácil de imprimir con equipo casero.		Fácil de imprimir con equipo casero.
Costo de impresión	Alto si se usa impresión profesional offset comparado con formas blanco y negro.			Bajo si se usa impresión profesional offset comparado con formas de segregación de color.		Bajo si se usa impresión profesional offset comparado con formas de segregación de color.
Tamaño de archivo		Las imágenes son pequeñas.	Las imágenes son grandes.		Las imágenes son grandes.	
Modo de escaneo	Se requiere de hardware/software con capacidad de filtros de color.		Rango muy limitado para el manejo del contraste y brillo.		Rango muy limitado para el manejo del contraste y brillo.	
Velocidad de procesamiento		Promedio de velocidad alto en procesamiento.	Promedio de tiempo de procesamiento alto, dado que requiere de destramado.		Errores frecuentes de procesamiento al encontrar caracteres pegados o traslapados con los bordes de campos.	
Calidad de reconocimiento		Calidad de reconocimiento muy alta.		Calidad de reconocimiento alta.	La calidad depende en gran medida de la nitidez del llenado de la forma.	Calidad de reconocimiento alta si es que la forma fue llenada correctamente.
Localización de textos explicativos		Pueden ser ubicados en cualquier lugar, inclusive dentro de los campos al imprimir con el color de fondo.	Requiere de mucho cuidado en cuanto a la ubicación de textos y otros elementos.		Requiere de mucho cuidado en cuanto a la ubicación de textos y otros elementos.	
Apariencia de la forma y legibilidad	El bajo contraste de fondo y los textos explicatorios pueden impactar negativamente en la legibilidad.	Muy estético, elegante y atractivo.	Baja legibilidad. Menos atractivo a la vista.		Menos atractivo a la vista.	Buena legibilidad.
Nitidez en el llenado		Menos sensible al traslapo de los textos con los bordes de campos.	Caracteres pequeños deben escribirse cuidadosamente y deben ser lo suficientemente grandes para no ser removidos como basura.	Menos sensible al traslapo de textos con los bordes de campos.	Muy sensible al traslapo de textos con los bordes de campos.	

FIGURA 1-23: FACTORES A CONSIDERAR EN EL DISEÑO DE FORMAS.

2. Análisis Situacional del Área de Talento Humano

2.1 Descripción, Importancia y Justificación de las Evaluaciones de Personal

Una prueba de evaluación para talento humano acompañada de un instrumento y el procedimiento adecuado, es la manera de obtener una medición objetiva y estandarizada de una muestra o grupo de candidatos, conocidos como sujetos, para determinar psicométricamente su comportamiento, evaluar su conocimiento, capacidades, habilidades y otras características de este sujeto con relación a otros.

El área de talento humano en el Ecuador no ha tenido la importancia que merece, encontrando que varias de las prácticas aplicadas a su gestión son anticuadas y anti técnicas. La evaluación de personal para su selección y medición de desempeño, así como los instrumentos utilizados como pruebas cognitivas, pruebas de aptitud, pruebas de rendimiento, inventarios de personalidad e intereses, pruebas de conocimiento del puesto a aplicar y pruebas de casos pueden carecer de características psicométricas mínimas recomendadas para su interpretación, pudiendo aportar a la conclusión de que dichos procesos evaluativos no aportan con información suficiente e importante respecto al nivel de desarrollo de los atributos evaluados.

Numerosas organizaciones privadas y multinacionales de mediano y gran tamaño, cuentan por lo general con psicólogos industriales de planta ó un departamento dedicado a la gestión del talento humano, cuyas funciones primarias, sin limitarse a las mencionadas, se orientan a seleccionar, contratar, entrenar o formar, emplear y retener a los colaboradores dentro de la organización, establecer políticas y procedimientos, administrar nómina, planes de compensación, relaciones contractuales y sindicalistas, promover el liderazgo y el trabajo en equipo, en conjunto y apoyados por los altos directivos de las mismas. En las organizaciones del sector público sucede generalmente lo contrario, ya que en la mayoría de casos no cuentan con psicólogos industriales de planta o un departamento para la gestión del talento humano, y recurren a servicios externos para apalancarse en la ejecución de varias de las tareas asociadas a la gestión del talento humano.

Entre las funciones descritas, el ciclo de reclutamiento, pruebas de selección y evaluaciones continuas de desempeño, probablemente es una de las funciones más críticas y demandantes. Su ejecución, que no siempre puede ser iniciada desde un computador, por diferentes razones tal como se expuso en la introducción del capítulo 1, se apoyan en su lugar en el método tradicional basado en lápiz y papel, lo que implica para el personal del área de talento humano tareas arduas de organización y clasificación de documentos, digitación y tabulación de datos, calificación, tratamiento estadístico y psicométrico para la posterior generación de informes y toma de decisiones, prueba por prueba y sujeto por sujeto, lo cual repercute de manera directa en la demanda de tiempo y recursos.

Si bien, una vez que los datos han sido ya poblados en una base de datos para su posterior tratamiento, usando una simple hoja de cálculo o un sistema específico, varias de las actividades siguientes a esta podrían ser automatizadas casi en su totalidad a partir de estas bases de datos. Pero, cómo llegaron estos datos desde el papel a la base de datos? Cuál fue el método que se utilizó para tabular los datos y cómo se aseguró la calidad de ingreso de los mismos a la vez que se garantizó que no se cometieron errores durante su ingreso?

Es un hecho que muchas organizaciones a pesar de contar con un departamento formal de talento humano el cual forma parte de su estructura funcional, se apoyan de consultores experimentados y de trayectoria para la medición de capital humano así como proveedores de servicios dedicados a la búsqueda de talentos con la finalidad de tercerizar justamente procesos de ejecución de pruebas de selección y evaluaciones de desempeño. Esta tercerización de servicios puntuales no genera conflicto respecto al Mandato Constituyente Número 8, el cual cita en su artículo número 1:

Se elimina y prohíbe la tercerización e intermediación laboral y cualquier forma de precarización de las relaciones de trabajo en las actividades a las que se dedique la empresa o empleador. La relación laboral será directa y bilateral entre trabajador y empleador. [C]

2.2 Normativa Legal para el Ámbito Público [D]

La expedición de la Norma del Subsistema de Reclutamiento y Selección de Personal amparada en los siguientes artículos de la LOSEP²⁴:

Artículo 54: *El sistema integrado de desarrollo del talento humano del servicio público estará conformado entre otros, por el subsistema de reclutamiento y selección de personal.*

Artículo 65: *El ingreso a un puesto público será efectuado mediante concurso de méritos y oposición el mismo que evalúe la idoneidad de los interesados y garantice el libre acceso a los mismos, considerando la inserción y accesibilidad en igualdad de condiciones al trabajo remunerado de las personas con discapacidad y de las comunidades, pueblos y nacionalidades, aplicándose acciones afirmativas para ese caso.*

Artículo 68; *El ascenso se realizará mediante concurso de méritos y oposición, en el que se evaluará, primordialmente, la eficiencia de las servidoras y servidores y, complementariamente, los años de servicio. A más de que, se deberán cumplir con los requisitos establecidos para el puesto,*

establece en los siguientes artículos:

Artículo 18.- De la verificación de postulaciones.- *Constituye el procedimiento mediante el cual, una vez cumplida la fecha máxima de entrega de postulaciones, la UATH²⁵ analiza y revisa las hojas de vida enviadas electrónicamente por las y los aspirantes y las convalida con los requisitos de la convocatoria, determinando su cumplimiento o no.*

La convalidación se registrará electrónicamente en el Formulario "Verificación de Postulaciones" (SEL-05) que, para conocimiento de las y los aspirantes, será publicado automáticamente a través la página web www.socioempleo.gob.ec,

²⁴ LOSEP: Ley Orgánica del Servicio Público.

²⁵ UATH: Unidad para la Administración del Talento Humano.

indicando la fecha, hora y lugar en la que las y los aspirantes aprobados deberán presentarse para rendir la pruebas de conocimientos y competencias técnicas.

Al finalizar este procedimiento, la UATH elaborará el acta de verificación de postulaciones, que será puesta en conocimiento del Tribunal de Méritos y Oposición.

Las y los aspirantes que, luego del análisis y verificación, cumplan los indicados requisitos, participarán en el procedimiento de evaluación. En el caso de no haber superado la verificación de postulaciones, podrán apelar electrónicamente, directamente a través de la página web www.socioempleo.gob.ec, en el término de dos días contados a partir de la publicación del formulario SEL-05, en la indicada página; para este efecto, la UATH al realizar esta publicación, hará constar obligatoriamente, la fecha y hora en la cual vencerá el término para presentar apelaciones.

El Tribunal de Apelaciones se pronunciará en el término de dos días, contados a partir de la fecha de vencimiento del término de apelaciones. En el caso de que las apelaciones presentadas para un mismo puesto sean superiores a cincuenta, se considerarán los siguientes días adicionales:

Numero De Apelaciones Presentadas	Término Adicional
Entre 50 a 100	1 día adicional
Entre 101 a 150	2 días adicionales
Entre 151 a 200	3 días adicionales
Más de 200	4 días adicionales

Una vez resueltas las apelaciones, el respectivo Tribunal de Apelaciones comunicará a la UATH el acta con los resultados, a fin de que esta los dé a conocer a las y los aspirantes, a través de la página www.socioempleo.gob.

Artículo 19.- De la evaluación.- *Concluida la verificación de postulaciones y resueltas las apelaciones, la UATH institucional iniciará el procedimiento de evaluación, mediante el cual se mide, objetiva y técnicamente, los niveles de competencias que ostentan las y los aspirantes a un puesto público, a través de*

pruebas de conocimientos y competencias técnicas, pruebas psicométricas y entrevistas.

La ponderación total de este procedimiento será sobre cien (100) puntos, asignados de la siguiente manera:

- Pruebas de conocimientos y competencias técnicas: Representarán 45 puntos del total.
- Pruebas psicométricas: Representarán 20 puntos del total; y,
- Entrevistas: Representarán 35 puntos del total.

Artículo 20.- De las pruebas de conocimientos y competencias técnicas.- *Estas pruebas evalúan el nivel de conocimientos y el grado de desarrollo de las competencias técnicas inherentes al perfil y la descripción del puesto requerido, que ostentan las y los aspirantes, incluirán aspectos de carácter práctico, para todos los aspirantes cuando el puesto lo amerite.*

Con los bancos de preguntas remitidos por las unidades, durante la preparación del proceso de reclutamiento, la UATH elaborará las pruebas correspondientes, las tomará a las y los aspirantes que superaron el procedimiento de verificación de postulaciones y apelaciones y las calificará. Posteriormente, registrará estos resultados electrónicamente en los formularios SEL-06 "Calificación de pruebas de conocimientos y competencias técnicas" y SEL-08 "Puntaje del componente de Evaluación y Puntaje Final".

En las entidades de la administración pública central, institucional y dependiente, las UATH institucionales remitirán el banco de preguntas elaborado al INSTITUTO NACIONAL DE LA MERITOCRACIA, para la conformación progresiva de un banco nacional de datos.

Artículo 21.- Puntaje mínimo para pruebas de conocimientos y competencias técnicas.- *Únicamente aquellos aspirantes que obtengan en las pruebas de conocimientos y competencias técnicas una calificación igual o superior a treinta sobre cuarenta y cinco (30/45) puntos, podrán continuar con la presentación de las pruebas psicométricas y las entrevistas.*

Estos resultados y el acta respectiva serán puestos en conocimiento del Tribunal de Méritos y Oposición.

Las y los aspirantes que no hubieran alcanzado el puntaje mínimo señalado, podrán apelar electrónicamente, a través de la página web www.socioempleo.gob.ec, en el término de dos días contados a partir de la publicación del formulario SEL-06, en la indicada página, para este efecto, la UATH al realizar esta publicación, hará constar obligatoriamente, la fecha y hora en la cual vencerá el término para presentar apelaciones.

El Tribunal de Apelaciones se pronunciará en el término de dos días, contados a partir de la fecha de vencimiento del término de apelaciones. En el caso de que las apelaciones presentadas para un mismo puesto sean superiores a cincuenta, se considerarán los siguientes días adicionales:

Numero De Apelaciones Presentadas	Término Adicional
Entre 50 a 100	1 día adicional
Entre 101 a 150	2 días adicionales
Entre 151 a 200	3 días adicionales
Más de 200	4 días adicionales

Una vez resueltas las apelaciones, el respectivo Tribunal de Apelaciones comunicará a la UATH el acta con los resultados, a fin de que esta los dé a conocer a las y los aspirantes, a través de la página web www.socioempleo.gob.ec.

Artículo 22.- De las pruebas psicométricas.- *La UATH aplicará las pruebas psicométricas a las y los aspirantes que alcanzaron un puntaje igual o mayor a treinta sobre cuarenta y cinco (30/45) puntos, en las pruebas de conocimientos y competencias técnicas. Las pruebas psicométricas se establecerán sobre la base de las competencias conductuales, valores y aptitudes necesarias para el puesto sujeto a concurso. Su calificación será sobre veinte (20) puntos y se registrará electrónicamente en el formulario SEL-08 "Puntaje del componente de Evaluación y Puntaje Final".*

Respecto a las pruebas psicométricas, la UATH deberá proveerse de un conjunto de baterías que hayan sido diseñadas en base a haremos ecuatorianos, para aplicarlas a las y los aspirantes en los concursos de selección.

Estas pruebas serán valoradas de acuerdo a las capacidades necesarias para cada puesto, y deberán ser aplicadas y calificadas por un profesional con conocimiento en la materia.

Para el efecto, las instituciones capacitarán a las o los servidores de la UATH cuya formación les permita realizar tales funciones; y, en caso de no disponer del personal adecuado, se podrá contratar los servicios profesionales de un psicólogo industrial experto durante la realización del concurso, exclusivamente para la aplicación y valoración de las pruebas psicométricas, de conformidad con la disponibilidad presupuestaria. El contenido de las pruebas psicométricas se mantendrá en estricta reserva hasta su aplicación.

Si bien esta normativa aplica al sector público, muchas organizaciones en el sector privado han aplicado varios de los procedimientos y pruebas contempladas en esta norma desde hace años atrás sin que esto implique un procesamiento de las evaluaciones de talento humano con captura electrónica de datos.

2.3 Procesamiento Manual de Evaluaciones de Talento Humano [4]

Los métodos tradicionales para procesar manualmente las evaluaciones de personal comprenden varias actividades demandantes de tiempo y control humano para poder ofrecer datos fidedignos posibles de ser usados para la generación de resultados psicométricos.

Una representación típica de las macro actividades que comprenden el procesamiento manual de evaluaciones de personal son representadas en la siguiente figura así como la descripción de las actividades:

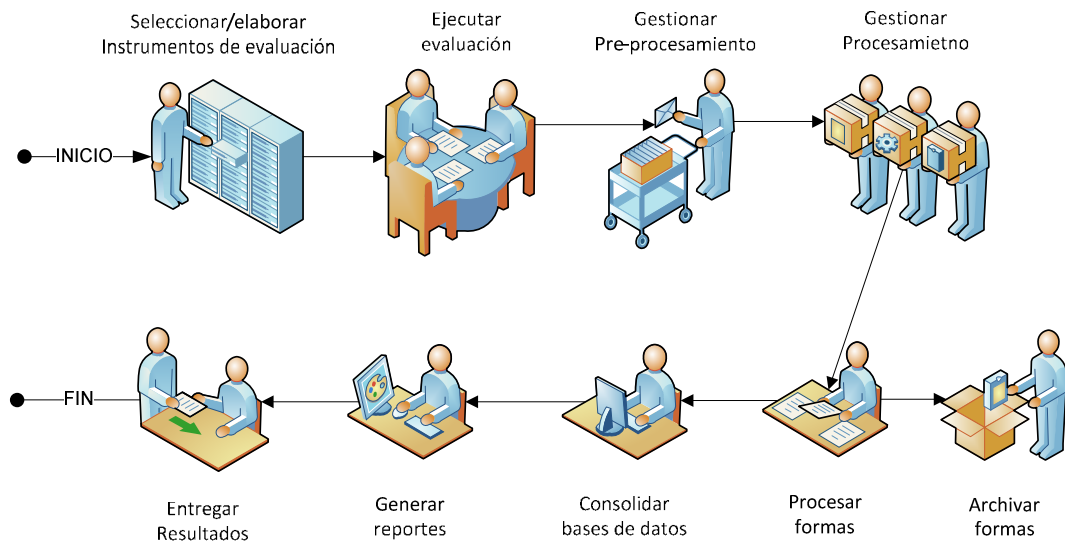


FIGURA 2-01: PROCESO MACRO PARA EL PROCESAMIENTO MANUAL DE EVALUACIONES DE PERSONAL. [A]

2.3.1 DESCRIPCIÓN DE ACTIVIDADES

Seleccionar/elaborar instrumentos de evaluación:

Selecciona los instrumentos de evaluación previamente creados por el evaluador o consultor que son posibles de aplicar ó en su defecto crea instrumentos nuevos de evaluación según la necesidad de aplicación o solicitud del cliente. Cada instrumento de evaluación comprende:

- Cuadernillos o banco de preguntas, con las instrucciones y todas las preguntas a ser resueltas por la persona evaluada.
- Batería de respuestas asociada al tipo de prueba ejecutada que contiene el juego de respuestas correctas con la que deberá calificarse a la persona evaluada.
- Formas para el registro de respuestas usado por la persona evaluada durante el desarrollo de la evaluación.

Ejecutar evaluación:

Las personas a evaluar reciben el cuadernillo de preguntas con una o más formas de respuestas según el tipo y número de pruebas a rendir, a la vez que reciben una inducción respecto al manejo del material entregado.

Al finalizar esta etapa, todo el material es recolectado, las formas de respuestas son clasificadas según el tipo de prueba y puestas en un sobre identificando información del lote, lugar u otros atributos requeridos. Es importante recalcar que la ejecución de la evaluación se la realiza en sitio donde el cliente lo requiera, pudiendo darse en varias localizaciones, de modo simultáneo e inclusive en diferentes ciudades.

Gestionar pre-procesamiento:

Los sobres cerrados son enviados a la central de procesamiento. Una vez recibidos en la central son registrados y repartidos a las diferentes personas (procesadores) para iniciar con el procesamiento.

Gestionar procesamiento:

Cada procesador abre un sobre (lote) y procede a identificar el tipo de prueba para seleccionar la batería de respuestas que deberá usar para proceder con la calificación. Adicionalmente prepara la hoja de cálculo para el registro de puntuaciones directas.

Procesar formas:

Para cada una de las pruebas del lote, el procesador realiza la calificación a cada pregunta poniendo una marca en aquellas cuya respuesta seleccionada coincide con la indicada en la batería de respuestas, es decir, la respuesta acertada. Finalizada la calificación de una prueba realiza la sumatoria de respuestas correctas y antes de registrar la puntuación directa en la hoja de cálculo, podría realizar una segunda verificación de las respuestas marcadas en la hoja contra la batería de respuestas. Esta actividad se repite tantas veces como pruebas contenga el lote.

Una vez finalizado el lote, las pruebas se guardan en el sobre para su archivo posterior.

Consolidar bases de datos:

Un usuario designado es el encargado de consolidar cada una de las hojas de cálculo individualmente creadas por cada procesador en la actividad anterior en una sola hoja de cálculo o base de datos.

Generar reportes:

De la base de datos consolidada se generan los siguientes reportes:

- Reporte con puntuaciones directas (por género u otras variables).
- Reporte con puntuaciones estandarizadas para cada sujeto y prueba.
- Análisis por ítems para la determinación de bondad de prueba, dificultad y discriminación de grupos buenos y malos de evaluados.
- Análisis psicométricos.

Entregar resultados:

Comprende la entrega de los reportes psicométricos al cliente y las formas de respuesta llenas y calificadas si es que son solicitadas. En caso de que el cliente no las solicita, las hojas de respuestas calificadas son archivadas como respaldo a procesos de reclamación.

Las actividades que podrían ser reemplazadas al implementar captura electrónica de datos van desde la gestión de pre procesamiento hasta la consolidación de base de datos, e inclusive la generación de formatos de datos listos para ser consumidos por sistemas de especializados de procesamiento psicométrico para la generación de estadísticas y reportes.

2.4 Debilidades Implícitas en el Procesamiento Manual

2.4.1 FORMAS DE RESPUESTAS

Debido a que las formas de respuestas, parten de un instrumento de evaluación respondiendo a una necesidad específica relacionada a un banco de preguntas, su generación es asistida usando un procesador de texto ó una hoja de cálculo y probablemente podría ser generada por más de una persona cuando se tratan de algunos instrumentos de evaluación. Al no usar una herramienta formal para la generación de formas, pueden presentarse como diseños no unificados y no estandarizados entre los diferentes instrumentos con lo que las instrucciones de llenado en cada banco de preguntas también no será estándar y unificada generando riesgo de que los datos proporcionados por la persona evaluada no estén completos, sean ilegibles,

inconsistentes o que generen retrasos en la interpretación al momento de ser procesados dado que el evaluado tiene la presión del tiempo para completar.

Adicionalmente, las formas de respuestas podrían no lucir elegantes ni profesionales si no son diseñadas con minuciosidad.

2.4.2 TIEMPO DE PROCESAMIENTO

El tiempo requerido para la revisión, calificación y registro de la puntuación directa en la hoja de cálculo en una prueba típica de 40 preguntas toma aproximadamente 4 minutos para un procesador que ya ha sido entrenado y tiene conocimiento respecto a las actividades asignadas. En casos en donde el número de preguntas es mayor, como en pruebas que alcanzan las 100 preguntas, el tiempo de procesamiento de cada una puede bordear los 7 minutos, descartando una relación directa entre el número de preguntas y el tiempo requerido para procesamiento.[4]

El escenario manual mostró que el tiempo para calificar, verificar y registrar la puntuación directa en la hoja de cálculo para un proyecto común con 4 instrumentos de evaluación y 40 preguntas cada uno ejecutado a 100 sujetos, es decir, un total de 400 evaluaciones, fue en promedio 5 minutos aproximadamente. Cabe indicar que este tiempo no considera los requeridos para tareas adicionales como el cierre de una hoja de cálculo previa a la consolidación ni el tiempo de transferencia entre lotes, así como la disponibilidad del procesador. [A]

La tabla siguiente muestra los parámetros y las métricas para el procesamiento manual de 400 evaluaciones con 1 procesador:

Parámetros de Procesamiento		
Descripción	Valor	Unidad
Instrumentos (lote)	4	na
Cantidad de evaluaciones por lote	100	na
Tiempo de proceso por prueba	5	minutos
Tiempo de preparación de lote	8	minutos
Tiempo para cierre de hoja de cálculo	10	minutos
Disponibilidad del procesador	90%	

FIGURA 2-02: PARÁMETROS DE PROCESAMIENTO PARA 100 SUJETOS CON 4 TIPOS DE PRUEBAS.

Métricas de Procesamiento (1 procesador)		
Descripción	Valor	Unidad
Número de procesadores	1	personas
Disponibilidad del procesador	90%	
Procesamiento de cada lote	555.56	minutos
Procesamiento de todos los lotes	2222.22	minutos
Preparación de todos los lotes	32	minutos
Cierre de hoja de cálculo de todos los lotes	3.6	minutos
Tiempo total de proceso	2257.82	minutos
Tiempo horas/hombre	37.63	horas
Días laborables requeridos	4.71	días

FIGURA 2-03: MÉTRICAS DE PROCESAMIENTO MANUAL PROYECTADAS CON UN PROCESADOR.

El tiempo consumido para las actividades de gestión y procesamiento de formas, son sin lugar los más críticos y sensibles para dar lugar a la introducción de errores. Así mismo la verificación de los datos introducidos por los procesadores son verificados por humanos con lo que no se puede dar una garantía de exactitud de datos al ciento por ciento.

2.4.3 DIFICULTAD PARA ESCALAR

Conocido el hecho de que un procesador trabaja con un solo lote a la vez para un proyecto específico para así reducir la variabilidad en su proceso y minimizar los errores, resulta casi imposible poder producir más con los mismos recursos si no se aumenta el número de personas para incrementar la producción respecto a un mismo proyecto o respecto a varios proyectos simultáneos en tiempos prudentes con diferentes instrumentos de evaluación.

La dificultad para aumentar la capacidad de producción tanto para un único proyecto de gran volumen o varios proyectos de mediano volumen, no es inmediatamente solucionada con la simple inclusión de recursos adicionales contratados para el fin específico, dada la curva de aprendizaje inherente a las actividades del proceso. Por otro lado, mantener recursos de procesamiento en nómina sin que sean productivos debido a que los ciclos de evaluación no son continuos, no resulta de manera alguna ser eficiente y rentable.

El proceso global desde la gestión de procesamiento hasta la entrega de resultados al cliente usando la metodología manual para un conjunto de 400 evaluaciones puede llegar a tomar entre 2 a 3 semanas lo que impacta de manera negativa en servicio al cliente y

consume altos recursos por la parte de quien ofrece el procesamiento de las evaluaciones.[4]

2.4.4 ERRORES EN INGRESO DE DATOS

Dado que el procesamiento es netamente manual, los errores en el ingreso de datos son causados por fatiga mental y visual, desconcentración y disminución de la atención asociadas a la ejecución continua de actividades iterativas que tienden a generar errores absurdos, los que resultan parecer complicados de encontrarlos cuando el operador está cansado. Un procesador debe necesariamente contar con lapsos de tiempo para descanso mental y visual con la finalidad que su trabajo mecánico pueda garantizar datos de calidad, aunque con esto se afecta la disponibilidad en su jornada laboral.

2.4.5 ARCHIVO DE EVALUACIONES

Las evaluaciones una vez procesadas son archivadas en los sobres de origen con la información del lote. En casos en donde se habría detectado un error, la única manera es recurrir al archivo físico y buscar la evaluación una a una dentro del lote.

Al tratarse de un archivo con cientos de pruebas podría no ser tan compleja la búsqueda, pero si se tratan de procesos en donde se evalúen a miles de sujetos, con 3 o 4 instrumentos, la búsqueda sería un reto si no se cuenta con un esquema eficiente para la organización del archivo físico.

2.4.6 RESPECTO A LA NORMATIVA LEGAL PARA EL ÁMBITO PÚBLICO

En las organizaciones públicas, el número de sujetos a evaluar podría crecer de manera considerable desde algunos cientos en instituciones locales hasta algunos miles a nivel nacional, en donde cada sujeto al menos rinde 3 diferentes tipos de pruebas.

Al atenerse a un marco de tiempo para el cumplimiento de los procesos de selección de personal, la clasificación y calificación manual de evaluaciones demandarán un gran espacio de tiempo para la generación de las bases de datos, junto a la posibilidad de la introducción de errores como se mencionó anteriormente.

La figura siguiente presenta el tiempo estimado de procesamiento manual con 5 procesadores para 2,500 sujetos, cada uno rindiendo 4 pruebas:

Parámetros de Procesamiento		
Descripción	Valor	Unidad
Instrumentos (lote)	4	na
Cantidad de evaluaciones por lote	2500	na
Tiempo de proceso por prueba	5	minutos
Tiempo de preparación de lote	8	minutos
Tiempo para cierre de hoja de cálculo	10	minutos
Disponibilidad del procesador	90%	

FIGURA 2-04: PARÁMETROS DE PROCESAMIENTO PARA 2,500 SUJETOS Y 4 INSTRUMENTOS.

Métricas de Procesamiento (3 procesadores)		
Descripción	Valor	Unidad
Número de procesadores	3	personas
Disponibilidad del procesador	90%	
Procesamiento de cada lote	13888.89	minutos
Procesamiento de todos los lotes	18518.52	minutos
Preparación de todos los lotes	10.67	minutos
Cierre de hoja de cálculo de todos los lotes	1.20	minutos
Tiempo total de proceso	18530.39	minutos
Tiempo horas/hombre	308.84	horas
Días laborables requeridos	38.61	días

FIGURA 2-05: MÉTRICAS DE PROCESAMIENTO MANUAL PROYECTADAS CON 3 PROCESADORES

Métricas de Procesamiento (5 procesadores)		
Descripción	Valor	Unidad
Número de procesadores	5	personas
Disponibilidad del procesador	90%	
Procesamiento de cada lote	13888.89	minutos
Procesamiento de todos los lotes	11111.11	minutos
Preparación de todos los lotes	6.40	minutos
Cierre de hoja de cálculo de todos los lotes	0.72	minutos
Tiempo total de proceso	11118.23	minutos
Tiempo horas/hombre	185.30	horas
Días laborables requeridos	23.17	días

FIGURA 2-06: MÉTRICAS DE PROCESAMIENTO MANUAL PROYECTADAS CON 5 PROCESADORES.

Adicionalmente, según la normativa, los postulantes legalmente pueden apelar en caso de sentirse perjudicados por su puntuación. Dado que el procesamiento de las evaluaciones fue realizado manualmente, no se podría garantizar la confiabilidad de los datos a un ciento por ciento, además de que se dependería de un archivo físico voluminoso en el cual se buscarían los documentos de origen para concluir dicha apelación.

2.5 El Costo del Procesamiento Manual a Mayor Escala

El cuadro a continuación presenta un resumen del cálculo con las variables asociadas al escenario de procesamiento manual para captura de datos a mayor escala, el cual considera los siguientes parámetros iniciales:

1. Evaluaciones continuas mensuales a 2,500 sujetos con 4 instrumentos distintos, es decir, 10,000 evaluaciones mensuales.
2. Salario básico más beneficios de ley para cada procesador.
3. En promedio las evaluaciones contienen 115 caracteres, el cual viene calculado a partir de:
 - a) El número de preguntas para cada tipo de evaluación:
 - i) Evaluación 1 con 40 preguntas.
 - ii) Evaluación 2 con 71 preguntas.
 - iii) Evaluación 3 con 86 preguntas.
 - iv) Evaluación 4 con 105 preguntas.
 - b) 10 dígitos para la cédula para cada evaluación.
 - c) 15 caracteres en promedio para los nombres para cada evaluación.
 - d) 15 caracteres en promedio para los apellidos para cada evaluación.
4. El tiempo promedio para procesar cada evaluación, calificar, verificar y registrar la puntuación directa de cada evaluación es de 7 minutos.

Item	Variable	Valor	Procedencia
A	Días en el año	365	Datos
B	Cantidad de evaluaciones al año	120,000	
C	Días laborables al año (con 10 días feriados)	254	
D	Horas laborables al día	8.0	
E	Costo mensual por operador (incluido beneficios de ley)	\$ 355.00	
F	Promedio de caracteres por evaluación	115	
G	Tiempo de proceso por evaluación (minutos)	7.0	
H	Promedio de páginas a producir por día	472	B/C
I	Caracteres a ingresar por día	54,331	H*F
J	Cantidad de evaluaciones procesadas por hora	8.6	60/G
K	Velocidad media de entrada de datos por hora	986	F*J
L	Horas requeridas para entrada manual de datos	55.1	I/K
M	Operadores requeridos para entrada de datos	7	L/D
N	Costo laboral por operador de entrada de datos al año	\$ 4,260	E*12
O	Costo laboral total de operadores de entrada de datos al año	\$ 29,820	N*M

FIGURA 2-07: ESCENARIO DE COSTOS PARA PROCESAMIENTO MANUAL A MAYOR ESCALA.

De los resultados se obtiene que se requiere de al menos 7 operadores de procesamiento para poder capturar un volumen de 473 evaluaciones de una página cada una por día con un promedio de 115 caracteres por cada evaluación bajo las condiciones de procesamiento antes descritas.

El costo anual proyectado por pago de salarios alcanza los US\$ 29,820 sin considerar recursos adicionales para otras actividades que se describen en el proceso manual de captura así como tampoco otros factores de costos como espacio de operación, mobiliario y equipos de cómputo.

3. Soluciones Propietarias y Libres para Captura Electrónica de Datos

De las alternativas listadas para captura electrónica de datos por lotes y captura electrónica de datos transaccional en el capítulo 1, el autor finalmente filtra a tres soluciones propietarias para ser analizadas en el presente capítulo, basándose en fundamentos netamente comerciales y de oportunidad de negocios para lograr calificarse como el primer representante y proveedor autorizado y certificado de soluciones y servicios de la marca a nivel país, y consecuentemente poder generar beneficios económicos sustentables contando con un respaldo global por parte del fabricante respecto a soporte y posicionamiento de marca.

En cuanto a soluciones libres para captura electrónica de datos, se encontró ciertamente alternativas amparadas en la licencia GPL²⁶, que de las encontradas, la mayoría ofrece alguno de los servicios de reconocimiento con OCR, ICR, OMR ó BCR, sin haber identificado alguna que combine todas las tecnologías de reconocimiento en una sola solución como lo hacen la mayoría de las soluciones propietarias.

Para cada una de las soluciones seleccionadas, se detallan características funcionales para finalmente poder elegir la que por sus fortalezas y beneficios en conjunto sería implementada como solución de captura electrónica de datos, principalmente para el procesamiento de evaluaciones aplicadas al talento humano y con la posibilidad de ofrecer otros servicios de captura de datos y documentos usando de preferencia la misma solución.

La información con la que se desarrolla este capítulo ha sido obtenida de manera directa desde el sitio en internet de cada desarrollador o fabricante de la solución, así como la información proporcionada a través de teleconferencias y demostraciones realizadas por medio de videoconferencia con algún representante comercial de las marcas con las cuales se pudo llevar a cabo estas sesiones. También se complementa con información y

²⁶ GPL: General Public License.

publicaciones obtenidas a través del portal de la Asociación para la Administración de la Información e Imágenes AIIM, quienes de manera independiente generan estudios de opinión respecto a las tecnologías de captura electrónica de datos implementadas en las organizaciones que conforman su comunidad de afiliados.

Los factores que se buscan analizar a medida de la posibilidad de contar con acceso a la información que cada fabricante o desarrollador presente respecto para su respectiva solución son:

- Tecnologías de captura y reconocimiento soportadas.
- Escalabilidad.
- Funcionalidad.
- Herramientas incluidas.
- Métodos de integración.
- Casos de éxito.
- Licenciamiento y costo.
- Entrenamiento y certificación.
- Soporte y mantenimiento.

3.1 Soluciones Libres y de Código Abierto

Si bien existen en el mercado soluciones de captura electrónica de datos bajo el esquema de licenciamiento GPL, las cuales permitirían el procesamiento de evaluaciones de talento humano con el uso de tecnología de reconocimiento óptico de marcas OMR y reconocimiento inteligente de caracteres ICR, se puede evidenciar en sus sitios de internet, que ninguna de las evaluadas muestra antecedentes respecto a su madurez ó casos de éxitos que prueben su aplicación en proyectos de dicha índole.

Acerca de la licencia GPL [5]

Es un tipo de acuerdo de licencia creada por la FSF²⁷ en 1,989, la cual se destina a la protección de la libre distribución, modificación y uso del software con la intención

²⁷ FSF: Free Software Foundation.

de evitar que los programas amparados bajo esta licencia puedan ser apropiados y restrinjan a los usuarios de las libertades que este acuerdo engloba.

Los fundamentos de libertad que la licencia GPL protege y que determinan para que un software sea identificado como software libre son: [6]

1. La libertad de usar el software para cualquier propósito (Libertad 0).
2. La libertad para modificar el software de acuerdo a las necesidades del usuario (libertad 1).
3. La libertad para compartir el software con el prójimo (libertad 3).
4. La libertad para compartir los cambios realizados en el software (libertad 4).

Las libertades 1 y 3 mencionadas requieren obligatoriamente contar con el código fuente del software, lo que quiere decir que el software libre es necesariamente de código abierto.

Cabe recalcar que el software gratuito no necesariamente tiene que ser software libre o de código abierto. El software de código abierto se enmarca en 10 premisas para poder ser identificado como tal:

1. Distribución libre: Debe ser regalado o vendido de manera libre.
2. Código fuente: Debe estar incluido o debe poder obtenerse de manera libre.
3. Derivación de trabajos: La redistribución de modificaciones debe estar permitida.
4. Integridad del código fuente del autor: Las licencias pueden requerir que las modificaciones sean redistribuidas como actualizaciones o parches.
5. No discriminación de personas o grupos: Nadie puede quedarse fuera.
6. No discriminación de áreas de iniciativa: Usuarios comerciales no pueden ser excluidos.
7. Distribución de licencia: Se aplican los mismos derechos a todos los que reciban el software.
8. La licencia no es específica para un producto: El programa no debe licenciarse solo como parte de una distribución mayor.

9. La licencia no restringe otro software: No obliga a que algún otro software que se distribuye con el de código abierto sea también de código abierto.

10. Licencia tecnológicamente neutral: No debe requerirse la aceptación de la licencia por medio de un clic de ratón o de otra forma específica del medio de soporte del software.

Los grandes proyectos de software basados en este acuerdo de licencia subsisten por el patrocinio de empresas que creen y apoyan la iniciativa de producto convirtiéndose en sus patrocinadores directos. En proyectos menores, se apalancan económicamente por las donaciones voluntarias realizadas por usuarios que están utilizando el producto así como también por el trabajo de cientos o miles de usuarios alrededor del mundo que están colaborando para mejorar el software.

3.1.1 *queXF* VERSIÓN 1.13.0 [7]

Desarrollada por la ACSPRI²⁸ como una solución de código abierto, basada en web para la verificación de formas de papel y entrada de datos, para eliminar la necesidad de ingreso manual de datos reduciendo así la fatiga y el error. La primera versión fue lanzada a finales del año 2,008.

La suite *queXF* trabaja con tecnología OMR y posteriormente a haber sido entrenado también puede aplicar tecnología ICR para interpretar caracteres manuscritos. Un verificador mediante un navegador web podrá confirmar que los campos reconocidos se hayan poblado correctamente pudiendo aceptar o corregir cualquier dato. La suite *queX* se basa en otros programas de código abierto para su funcionamiento, los que comprenden:

- **queXML**: Descriptores XML²⁹ con hojas de estilos asociadas para producir cuestionarios electrónicos (LimeSurvey³⁰) y cuestionarios basados en papel.

²⁸ ACSPRI: Australian Consortium for Social and Political Research Incorporated.

²⁹ XML: eXtensible Markup Language.

³⁰ LimeSurvey es una aplicación de código abierto para manejo de encuestas electrónicas. [8]

- **queXF**: Sistema basado en web para el escaneo y verificación de cuestionarios producidos con queXML.
- **queXS**: Sistema de código abierto para asistencia de entrevista telefónica basada por computador.
- **queXC**: Sistema de código abierto basado en web para limpieza y codificación de datos.

Desde la versión 1.12, queXF cuenta con tecnología ICR para la detección de caracteres aislados o independientes manuscritos, sin embargo el sistema aún está en pruebas y podría correr de manera lenta. El sistema ICR de queXF requiere de entrenamiento a partir de juegos de caracteres encontrados en una forma en uso o a través de la importación de juegos de caracteres externos.

Funcionalidades

- Número ilimitado de operadores de verificación.
- Alta precisión en la detección de marcas de selección.
- Toda la operación es basada en web, por lo tanto el procesamiento puede ser distribuido.
- Posibilidad de visualizar las formas digitalizadas vía web para poder comprobar que los datos son exactos.
- Detección automática de la localización de marcos de selección sobre la forma.
- Posibilidad de implementar ICR sobre la forma para el reconocimiento de texto y números.

Foro de usuarios y soporte

El foro de la comunidad de queXF, que registra a la fecha 45 usuarios, muestra actividad desde marzo de 2,008 hasta la noviembre de 2011, con entradas de soporte, noticias y actualizaciones. La mayoría de estas entradas pertenecen a Adam Zammit quien figura como el desarrollador principal de la suite queX. Comercialmente, la ACSPRI ofrece paquetes con servicios para soporte, instalación, personalización y administración de la suite queX a un valor de US\$ 75.00 por hora.

3.1.2 SHARED QUESTIONNAIRE SYSTEM VERSIÓN 2.0 [9]

El desarrollo del proyecto SQS inicia como un sistema de soporte a las evaluaciones escolares en agosto de 2,003 cuya distribución de código fuente se publica en SourceForge.net en febrero de 2,004. SQS es un sistema integrado para procesamiento de formas con OMR con interfaces gráficas de usuario, orientada a desarrollar plataformas sociales para compartir conocimiento acerca de los cuestionarios basados en estándares XML, siendo un sistema de código abierto licenciado bajo la licencia Apache License versión 2.0, que puede ser usado, pirateado y redistribuido, aceptando contribuciones, parches y donaciones económicas.

Las aplicaciones SQS corren en ambientes multi plataforma que tengan instalado JRE³¹ versión 6 y un navegador web y consta de los siguientes componentes:

- **SourceEditor 2.0:** Editor XML que actúa como diseñador de formas de interface gráfica.
- **MarkReader 2.0:** Procesador de formas OMR que usa imágenes digitalizadas para extraer las respuestas.

Funcionalidades

- Las formas OMR son publicadas como archivos PDF usando el procesador embebido que incluye tipografías TrueType. Estos recursos están disponibles en idioma inglés y japonés.
- El procesamiento corre en un subsistema de procesos paralelos distribuidos sin requerimientos de configuración.
- Los resultados OMR pueden ser navegados y editados usando interfaces AJAX³²
- Los operadores pueden ingresar respuestas a preguntas de formato abierto de manera manual.

³¹ JRE6: Java Runtime Environment.

³² AJAX: Asynchronous JavaScript And XML.

Foro de usuarios y soporte

El foro de soporte de SQS muestra actividad desde diciembre de 2,009 hasta marzo de 2010 para la versión SQS 2.0 y con 4 entradas adicionales a partir de septiembre de 2,009 hasta julio de 2011 para la versión en prueba SQS 2.1. El único desarrollador de SQS y administrador del foro registra ser Hiroya Kubo.

3.1.3 UDAI OMR TOOL [10]

Es una herramienta OMR libre destinada para uso en organizaciones sociales sin fines de lucro para realizar encuestas y tabular datos marcados en formas de papel, diseñada especialmente para extraer datos desde formas que han sido fotocopiadas y escaneadas.

Este proyecto fue iniciado por Aniket Kate and Maheedhar Kolla y desarrollado en su totalidad por Aaditeshwar Seth del *Udai Waterloo Chapter*, de la universidad de Waterloo. Udai es una organización de estudiantes voluntarios presente en la universidad de Waterloo en Canadá y de la universidad de California en San Diego.

Funcionalidades

- Para la creación de las formas usa una plantilla de Microsoft® PowerPoint® la que contiene dos círculos concéntricos en la esquina superior izquierda e inferior derecha para que el software pueda controlar la alineación y el ángulo de rotación. Esta plantilla de ejemplo se descarga desde el sitio en internet del proyecto.
- Las formas pueden ser fotocopiadas sin problema alguno.
- La definición de los campos de selección que serían interpretados por el software son comandos netamente de texto que se los describe en un editor común.
- Udai OMR Tool no ha sido probado con imágenes de diferentes resoluciones, por lo tanto la plantilla y las formas digitalizadas deberían mantener la misma resolución.
- El código fuente aún no se encuentra documentado, sin embargo está en proceso.
- Los datos capturados de las formas aún no pueden ser puestos directamente en una base de datos.

Foro de Usuarios y Soporte

A la fecha no existe un foro de interacción entre usuarios, aunque existe disponible una dirección de correo electrónico para contactar en caso de inquietudes ó errores respecto al software.

3.2 Soluciones Propietarias

3.2.1 ABBYY FLEXICAPTURE VERSIÓN 10 [11]

Fundada en 1,989 como BIT Software en Moscú, Rusia. En septiembre de 1,997 cambia su nombre a ABBYY y es actualmente un proveedor líder de software internacionalmente posicionado en conversión de documentos, captura de datos y servicios lingüísticos basando su investigación y desarrollo en inteligencia artificial para el reconocimiento de imágenes. Hasta enero de 2,011 contaba con más de 1,000 empleados entre sus 9 oficinas ubicadas en Rusia, Alemania, Reino Unido, USA, Japón, Taiwan y Ucrania. Sus productos son usados en más de 130 países con un total aproximado de 30 millones sus usuarios.

ABBYY investiga y desarrolla su propia tecnología la cual es aceptada e incorporada como motor OCR por varios fabricantes de software para procesamiento de imágenes y reconocimiento de texto.

Entre la gama de productos desarrollados por ABBYY, en 2,007 lanza la primera versión de FlexiCapture versión 8 la que nació de su predecesor FormReader 6 ya discontinuado que procesaba solamente formas fijas. ABBYY FlexiCapture es una solución de software de última generación para captura y procesamiento de documentos altamente escalable, preciso e inteligente que provee un punto único de entrada para clasificar, separar y transformar documentos y formas de cualquier tipo, estructura y complejidad en datos listos para ser consumidos por sistemas de negocio. FlexiCapture 10 es lanzado comercialmente a mediados de 2,011.

ABBYY FlexiCapture está conformado por los siguientes módulos:

- **ABBYY FormDesigner:** Módulo para crear e imprimir formas fijas o estructuradas en blanco para ser posteriormente completadas a mano. Estas formas pueden ser

impresas a papel o publicadas en formatos PDF para ser enviadas y completadas por correo electrónico.

- **ABBYY FlexiLayout Studio:** Herramienta que permite la creación de plantillas para el procesamiento de formas flexibles (semi estructuradas o no estructuradas) y clasificadores documentales usados para la identificación y clasificación automática documental.
- **Document Definition Editor:** Módulo que permite proveer las configuraciones y reglas a las plantillas documentales fijas o flexibles creadas en alguno de los componentes antes descritos.
- **ABBYY FlexiCapture:** Es en sí la solución misma para la captura electrónica de datos.

Tecnologías de captura soportadas

ABBYY FlexiCapture trabaja con tecnologías OCR, ICR, OMR y BCR independiente de la edición y tipo de licenciamiento del producto, sobre formas fijas y flexibles que incluyen documentos de una sola página o de múltiples página con o sin tablas.

Funcionalidades

- Una única suite integral para cubrir las necesidades de procesamiento de formas fijas y flexibles que incluye sin restricción alguna el uso de todas las tecnologías de reconocimiento.
- Fácil y rápida instalación y configuración de servidores y estaciones.
- Estaciones múltiples para administración y monitoreo basado en web.
- Captura de documentos con escáneres controlados por interfaces TWAIN, ISIS y WIA, carpetas de red con monitoreo automático o manual, ubicaciones FTP³³ y correo electrónico.
- Estaciones de digitalización remotas con opción de auto instalación.
- Estaciones de verificación tipo cliente servidor y basadas en web funcionales en 3 modos: Verificación en contexto de campo, verificación grupal y

³³ FTP: File Transfer Protocol.

verificación respecto a la imagen del documento, con opción de verificación de datos doble.

- Validación de datos contra fuentes externas de datos.
- Tecnología de tolerancia a fallas con Microsoft® Cluster.
- Clasificador multinivel inteligente basado en imágenes con tecnología de aprendizaje automático para proceso de clasificación de documentos y extracción de datos.
- Estaciones de procesamiento indefinidas en múltiples servidores o estaciones de trabajo.
- OCR para 180 idiomas, ICR para 100 idiomas y soporte de diccionario para más de 40 idiomas.

Integración

ABBYY FlexiCapture es configurable para exportar datos hacia formatos de archivo de texto plano, archivos de Microsoft® Excel®, bases de datos compatibles con ODBC³⁴, archivos XML y Microsoft Office SharePoint® Server.

Cuenta con la posibilidad de integrarse a sistemas de negocio ERP³⁵, DMS y ECM mediante scripts personalizados tanto para la exportación de datos e imágenes, además de contar con API³⁶ de servicios Web para integrarlo con distintas aplicaciones externas usando SOAP³⁷.

Otros métodos de integración a aplicaciones de terceros comprenden la creación de estructuras XML y programación con .Net Framework.

Casos de éxito

El sitio en internet de ABBYY contiene varias entradas con relatos de casos de éxito en diferentes industrias como gobierno, banca, seguros, educación, salud y más, en

³⁴ ODBC: Open Database Connectivity.

³⁵ ERP: Enterprise Resource Planning.

³⁶ API: Application Programming Interface.

³⁷ SOAP: Simple Object Access Protocol.

diferentes países alrededor del mundo como Rusia, India, Alemania, Singapur, Estados Unidos, Canadá, Chile, Reino Unido y Brasil entre varios más.

Proyectos notables que denotan la aplicabilidad a gran escala de esta solución son el procesamiento de las elecciones presidenciales en Chile en 2,009 que constó de 160,000 formas que se procesaron en 72 horas y el Ministerio de Educación en Rusia encargado de tomar pruebas estandarizadas de varias materias a más de 650,000 estudiantes a nivel nacional, lo que representó un aproximado de 3,3 millones de formas que fueron procesadas en 18 días.

Escalabilidad

Un mismo producto permite procesar volúmenes documentales bajos y medianos en configuración independiente así como volúmenes medianos y altos con una configuración distribuida cliente servidor. ABBYY FlexiCapture es un producto extremadamente escalable y flexible con el que de manera rápida y fácil se puede expandir los requerimientos crecientes de procesamiento en un mismo sitio ó en puntos remotos.

Licenciamiento y costos

ABBYY FlexiCapture puede licenciarse de manera perpetua y como licencia BPO. En el licenciamiento perpetuo el cliente adquiere un número de páginas a procesar potencialmente dentro de un año calendario, cuya cantidad de páginas en cada fecha de aniversario es renovada de manera automática sin tener que pagar nuevamente por la licencia. En el licenciamiento BPO el cliente adquiere un número de páginas fijo a ser procesados y dicha licencia no tiene fecha de caducidad sino al consumir el total de páginas adquiridas. En ambos casos el cliente paga el costo respectivo por concepto de soporte y mantenimiento cada año.

La licencia mínima para la instalación independiente que incluye 100,000 páginas al año para procesar formas fijas con una estación de verificación y una estación de digitalización bordea los US\$ 1,800, mientras que la misma configuración para formas fijas y flexibles se acerca a los US\$ 2,300.

La licencia mínima para la instalación distribuida que incluye 500,000 páginas al año para procesar formas fijas con dos estaciones de verificación y una estación de digitalización se acerca a los US\$ 7,500, mientras que la misma licencia para procesar formas fijas y flexibles se aproxima a los US\$ 10,500. Esta licencia incluye un número indeterminado de estaciones de procesamiento y de monitoreo.

Entrenamiento y certificación

ABBYY ofrece entrenamiento intensivo y certifica a sus clientes y nuevos socios de negocio con varios programas al año en diferentes ciudades en donde tienen presencia directa o un socio de negocios establecido.

Soporte y mantenimiento

El contrato de soporte y mantenimiento se suscribe con el pago del 20% del total de las licencias adquiridas por el cliente, obteniendo el derecho de solicitar y recibir soporte técnico de primer nivel a través del distribuidor certificado y autorizado de ABBYY ó de segundo nivel directamente de ABBYY Rusia mediante correo electrónico y teleconferencia. El mantenimiento le permite dentro de este mismo período acceder sin costo adicional a las mejoras menores del producto así como a las actualizaciones de versión de ABBYY FlexiCapture.

Puntuación de la comunidad AIIM [12]

Con un total de 201 votos generados por usuarios pertenecientes a la comunidad AIIM, ABBYY FlexiCapture es calificado con un valor global de 3.8 sobre 5 puntos.

Otros productos

ABBYY no solamente desarrolla su producto FlexiCapture para satisfacer las necesidades de captura electrónica de datos de un mercado corporativo, sino que además cuenta con otros productos orientados a un mercado de usuario final que ofrece transformación de documentos a diversos formatos y también otros productos corporativos que ofrecen servicios de OCR a demanda, como son:

- **ABBYY FineReader:** Solución de OCR para estaciones de trabajo que permite convertir imágenes a formatos estáticos posibles de ser buscados por texto

completo como PDF o TIFF o formatos editables para ser usados en procesadores de texto como Microsoft® Word® ó Excel®.

- **ABBYY Recognition Server:** Solución de OCR basada en servidor orientada para la digitalización masiva de archivos y servicios desatendidos de conversión de imágenes a formatos de documentos indexados con posibilidad de búsqueda.
- **ABBYY PDF Transformer:** Solución para conversión de diversos formatos de documentos a PDF y viceversa.
- **ABBYY Lingvo:** Solución que proporciona servicios de traducción instantánea, gramática y diccionario de palabras y frases desde cualquier aplicación en uso.
- **ABBYY Software Development Kit:** Permite el desarrollo e integración de sistemas con tecnología de captura electrónica de datos y conversión de documentos a un alto nivel técnico. El SDK incluye librerías que permiten acceder a la funcionalidad central de tecnologías de reconocimiento mediante interfaces API.

3.2.2 KOFAX CAPTURE VERSIÓN 10 [13]

Fundada en 1,985 en Irvine, California y adquirida por Dicom Group en 1,999 cambiando su nombre a Kofax en 2,008. Es un proveedor líder de soluciones de captura empresarial que garantizan un rápido retorno de inversión en la automatización de procesos para la carga de información intensiva de manera precisa, rápida y económica. Cuenta con subsidiarias en 30 países agrupando a más de 1,200 empleados. Sus productos están instalados en más de 70 países sobre una base de aproximadamente 20,000 clientes.

Kofax es una plataforma poderosa que provee un proceso estándar y consistente para la captura de todos los tipos de documentos entrantes e información. Kofax Capture agiliza procesos empresariales recopilando documentos y formularios, transformándolos en información precisa y recuperable disponible para aplicaciones y bases de datos como formatos estructurados.

Según estudios de mercado realizados por Harvey Spencer Associates, una firma de análisis independiente, indica que Kofax alcanzó para el segmento de captura por lotes en el año 2010, una participación de mercado del 35%, frente a un 15% de su competidor más cercano no revelado y una participación del 17% en captura transaccional.

Kofax Capture junto a la suite de aplicaciones adicionales KTM³⁸ potencializa la transformación de documentos en información electrónica estructurada lista para ser entregada a procesos de negocios, automatizando la clasificación y separación de documentos y la extracción de datos desde documentos de cualquier tipo independiente de su contenido y formato, es decir, desde formas semi estructuradas y no estructuradas de una o más páginas.

Tecnologías de captura soportadas

Kofax Capture soporta las tecnologías de reconocimiento OCR, ICR, OMR y BCR para formas fijas. Kofax Capture al ser complementado con KTM puede aplicar las mismas tecnologías de reconocimiento con opciones de clasificación documental sobre formas flexibles y libres.

Los motores de reconocimiento que KTM usa para OCR son ABBYY FineReader 9.0 desarrollado por ABBYY Rusia y RecoStar 4.0 desarrollado por Open Text Document Technologies en Alemania.

Funcionalidades

- Interface de usuario tipo Microsoft® Office® que provee una rápida familiarización para el usuario.
- Despliegue y uso de la aplicación desde un navegador de internet.
- Paquetes de lenguajes múltiples los que son añadidos al momento de la instalación.
- Posibilidad de buscar y filtrar lotes y dividir lotes basados en los resultados de clasificación documental para enviar documentos específicos a diferentes procesos, usuarios u operaciones.
- Captura de documentos y datos virtualmente desde cualquier fuente: Escáner, dispositivos MFP, flujos de impresión, correo electrónico, fax, servicios web o carpetas.

³⁸ KTM: Kofax Transformation Modules.

Integración

Kofax Capture cuenta con más de 140 conectores listos para integrarse con aplicaciones en varias líneas de negocio, sistemas ECM, ERP, BPM y soluciones de flujo electrónico.

Kofax Capture usa scripts de liberación para conectarse de manera transparente a sistemas como IBM®, Oracle®, Microsoft®, Open Text®, Hyland® entre otros más. También puede exportar datos a cualquier base de datos compatible ODBC y archivos delimitados ASCII³⁹.

Casos de éxito

El sitio de Kofax en internet muestra varios casos de éxito con aplicaciones de Kofax Capture a diferentes procesos de negocio. Un proyecto de aplicación a evaluaciones estudiantiles es el realizado por el Ministerio de Educación Iraquí el cual automatiza el procesamiento de evaluaciones a cerca de 6 millones de estudiantes de educación primaria y secundaria. La emisión de certificados para dicho proceso pudo reducirse de 3 meses y 45 personas involucradas a hasta 4 días con 10 personas.

Escalabilidad

Kofax Capture es una aplicación modular que puede ser usada para satisfacer las necesidades de captura departamental y ser expandida para cumplir con requerimientos de procesamiento complejos de alto volumen y misión crítica en la organización.

Kofax Enterprise es la solución diseñada para escenarios distribuidos de alta disponibilidad con recuperación ante desastres con consolas remotas para administración que monitorean el desempeño del sistema y permiten manejar excepciones de manera rápida y oportuna.

³⁹ ASCII: American Standard Code for Information Interchange.

Licenciamiento y costos

Kofax Capture se licencia en modo perpetuo en 3 modalidades: Standard, Stand Alone y Enterprise.

Kofax Capture Stand Alone funciona en un solo equipo con un mínimo de 60,000 páginas al año, el cual aproximadamente cuesta US\$ 1,400.

La configuración mínima para Kofax Capture Standard requiere de una estación concurrente y una estación de escaneo por volumen para 300,000 páginas al año el cual bordea los US\$ 5,200.

Kofax Capture Enterprise requiere al menos de una estación concurrente y una estación de escaneo por volumen para 600,000 páginas anuales con un costo aproximado de US\$ 9,500.

Para obtener capacidades de procesamiento avanzadas, Kofax Capture requiere de KTM el cual licencia según la funcionalidad específica requerida de procesamiento como: clasificación de documentos y separación, extracción de índices, extracción de índices en cabeceras de facturas, extracción de tablas o líneas de ítems en facturas entre otros más. El costo de estos módulos KTM varía por dos factores: El número de páginas anuales de base adquiridos para Kofax Capture, y la funcionalidad específica requerida. Para ilustrar un ejemplo, el módulo KTM para extracción de 5 índices para 60,000 páginas anuales de base bordea los US\$ 7,500 mientras que para las 300,000 páginas anuales de base se acerca a US\$ 21,000.

Entrenamiento y certificación

Kofax ofrece programas de entrenamiento para toda su gama de productos tanto en casa matriz, mediante un kit de entrenamiento basado en computador y con entrenamiento en línea. Una vez completado dicho entrenamiento, el candidato debe rendir el examen de certificación para calificar como un distribuidor autorizado del producto seleccionado.

Soporte y mantenimiento

Kofax maneja un sistema de incidencias en línea cuyo tiempo de respuesta varía entre uno de los siguientes esquemas de soporte adquirido: Standard y soporte 24x7. Para el primer caso un pago del 20% del total de las licencias adquiridas activa el contrato de soporte, mientras que en el segundo caso un pago del 28% del total de las licencias adquiridas es requerido. Esta última opción de soporte aplica únicamente para contratos de licencias que superan los US\$ 100,000.

Puntuación de la comunidad AIIM [14]

Kofax Capture recibe una puntuación global de 3.7 puntos sobre 5 de acuerdo a la votación de 2251 usuarios de la comunidad AIIM. Kofax Transformation Modules también obtiene una calificación global de 3.7 puntos sobre 5 en la votación de 282 usuarios de la comunidad.

3.2.3 IRISCAPTURE PRO FOR FORMS VERSIÓN 8.5 [15]

IRIS⁴⁰ inicia en 1,987 por Pierre De Muelenaere como una derivación de su tesis doctoral en un laboratorio universitario especializado en reconocimiento óptico de caracteres en la Universidad Católica de Lovaina, Bélgica. Desde 1,992, IRIS ha desarrollado y ha provisto de productos y tecnología para reconocimiento inteligente de caracteres y administración electrónica de documentos. Actualmente cuenta con operaciones en Bélgica, Francia, Luxemburgo, Alemania, Dinamarca, Noruega, Holanda, Estados Unidos y China, comercializando sus soluciones en más de 90 países alrededor del mundo. IRIS es actualmente el desarrollador de su propia tecnología para reconocimiento.

IRISCapture Pro for Forms incluye los siguientes módulos:

- **Designer:** Módulo que permite crear y diseñar las plantillas electrónicas para los documentos de papel a ser procesados por la aplicación principal. Estas permiten que cada tipo de documento sea identificado para que de manera fácil y rápida los datos requeridos puedan ser ubicados y leídos. Estas plantillas también contienen

⁴⁰ IRIS: Image Recognition Integrated Systems.

las instrucciones necesarias para el pre procesamiento de documentos así como las propiedades y reglas de validación a ser usadas sobre los datos reconocidos.

- **Scan/Import:** Permite que los documentos sean escaneados o tomados desde un sistema de archivos. También soporta la adquisición de imágenes a alta velocidad usando IRIS PowerScan.
- **Form Identification:** Encargado de buscar la coincidencia de la imagen digitalizada con la plantilla electrónica.
- **Read:** Módulo automático que ejecuta las tareas de reconocimiento sobre los datos de interés indicados en la plantilla electrónica. Puede utilizar varios motores de reconocimiento así como la tecnología Voting para elegir el reconocimiento más fiable empleando más de un motor de reconocimiento.
- **Job Quality Control:** El módulo JQC se encarga de separar aquellos documentos cuya calidad de imagen es menor a la requerida.
- **Correction:** Permite la corrección de datos de manera interactiva para poder revisar que los caracteres manuscritos hayan sido reconocidos apropiadamente por el sistema usando cuatro herramientas para corrección.
- **Validación:** Ejecuta las reglas de validación establecidas en la plantilla electrónica sobre cada uno de los campos.
- **Exportación:** Permite la exportación de los datos e imágenes pudiendo clasificarlas según el tipo de forma.

Tecnologías de captura soportadas

IRISCapture Pro for Forms soporta tecnologías OCR, ICR, OMR y BCR tanto en formas fijas como flexibles. Para la tecnología ICR, IRIS utiliza su tecnología IRIS ICR Voting, la cual provee una alta exactitud de reconocimiento.

Funcionalidades

- Para la digitalización de documentos usa escáneres con interfaces TWAIN ó mediante la interface IRIS FastLink soporta todos los escáneres KODAK con software Capture Perfect en modo color o dual stream. Esta arquitectura de doble flujo permite procesar documentos escaneados en color o en modo dual a la vez (color y escala de gris ó color y blanco y negro).

- Procesamiento automático de formas flexibles en el mismo sistema.
- Herramientas de supervisión que ayudan a evitar problemas operacionales durante el procesamiento. El módulo de auditoría ofrece una amplia variedad de reportes con indicadores gráficos y estadísticos para obtener control total sobre las actividades de usuarios y la productividad del sistema.
- Cuenta con una gran combinación de herramientas intuitivas y optimizadas para una rápida verificación y corrección de datos: Verificación masiva por lote, verificación en contexto, captura de datos en campos libres y codificación asistida por computador para campos vinculados.
- Arquitectura flexible basada en cliente servidor que funciona sobre sistemas basados en Windows®, Unix y Linux con diferentes bases de datos.
- Aplicación abierta que ofrece facilidad de integración de interfaces con otras aplicaciones y posibilita personalizar la aplicación usando API.
- ICR optimizado gracias a la tecnología ICR Voting de IRIS.

Integración

IRISCapture Pro for Forms puede integrarse con fuentes de datos mediante ODBC, acceso DLL⁴¹, API y línea de comandos. La exportación de datos es realizada a formatos XML y CSV.

Casos de éxito

El sitio en internet de IRIS contiene a manera de referencias solamente una descripción rápida de proyectos en donde se ha empleado IRISCapture for Forms y algunos nombres de clientes que usan IRIS la mayoría en países europeos.

Entre las referencias destacan: La digitalización y procesamiento de 20 millones de formas para el censo de población realizado por la oficina de estadística checa en 2,001, el cual se lo finalizó en un período menor de 5 meses; el procesamiento de formas flexibles de impuestos para el Servicio Público Federal Financiero en Bélgica con lo que se logró la digitalización y archivo de diferentes tipos de formas

⁴¹ DLL: Dynamic Link Library.

aumentando la producción de 60 a más de 600 formas por hora; y, el censo socio económico realizado por el Instituto Nacional de Estadísticas de Bélgica en el año 2,002 que contó con 15 millones de formas A3 a doble cara con un promedio diario de procesamiento de 150,000 formas.

Licenciamiento y costos

No se encuentra información adicional respecto a la forma de licenciamiento y costos de IRISCapture Pro for Forms en el sitio de internet de la compañía, sin embargo, se procedió a completar el formulario de registro de contacto y de registro de solicitud de socio de negocios requiriendo información al respecto pero aún no se ha obtenido respuesta hasta el momento.

Puntuación de la comunidad AIIM [16]

Con un total de 18 votos generados por usuarios pertenecientes a la comunidad AIIM, IRIS CapturePro es calificado con un valor global de 4.2 sobre 5 puntos.

Otros productos

IRIS fabrica hardware para consumo tales como escáneres tipo lápiz, móviles y para tarjetas de presentación, plumas digitales y software para conversión de documentos en formatos editables.

En su gama corporativa de productos y soluciones ofrece:

- **IRIS Document Server:** Una solución profesional y competitiva que mediante la provisión de servicio de OCR a múltiples usuarios permite la conversión de cualquier tipo de imagen en formatos editables e indexados.
- **IRIS PowerScan:** Solución para digitalización, estructuración, clasificación e indexación de documentos ideal para ser usada con escáneres de alta velocidad.
- **IRIS Toolkit & SDK:** Soluciones que permiten incluir las funcionalidades de los productos IRIS en aplicaciones propias de negocio ó crear aplicaciones propias basada en las tecnologías de reconocimiento IRIS.
- **IRISCapture Pro for Invoices:** Solución para digitalización, identificación, archivo y codificación de facturas.

- **IRISCapture Pro Sort & Index:** Solución para clasificación, indexación y archivo de documentos.

3.3 Selección de la Solución EDC

3.3.1 SOLUCIONES LIBRES Y DE CÓDIGO ABIERTO

Dadas las libertades que el acuerdo de licencia GPL comprende, resulta ser que el código fuente que gobierna el funcionamiento de una aplicación libre, puede ser modificado por cualquier persona alrededor del mundo. En casos en donde un cliente ha adoptado alguna mejora desarrollada por un tercero, ó en su defecto ha desarrollado e implementado por sí mismo una mejora, la responsabilidad de funcionamiento de la solución global recae prácticamente sobre estos últimos actores dejando sin responsabilidad de soporte al desarrollador primario de la aplicación.

De las soluciones analizadas para OMR con licencia GPL, además de que ninguna ofrece entrenamiento y certificación formal en el producto, el soporte técnico y el acceso a las mejoras del software depende casi en su totalidad de los aportes realizados por los mismos usuarios pertenecientes a la comunidad de cada proyecto y con la opción de contactar al desarrollador primario mediante correo electrónico. Esto causa que el tiempo de ciclo para corregir errores, generar parches y lanzar versiones nuevas con mejores características sea extenso y que prácticamente la posibilidad de obtener soporte formal sea muy limitado. Adicionalmente, el desarrollador primario de la aplicación, pierde el control sobre los cambios realizados en su producto, y no podría responsabilizarse por ofrecer soporte a los usuarios que utilizan el software con modificaciones o adiciones realizadas por terceros.

Este hecho no es el común denominador en software con licenciamiento GPL. Existen en el mercado sistemas operativos para servidores y estaciones de trabajo libres y de código abierto como *Red Hat* y *Ubuntu*. Estos gozan de gran aceptación y confiabilidad por parte de grandes organizaciones a nivel mundial e inclusive cuentan con certificaciones de funcionamiento por parte de prestigiosos fabricantes de hardware como Sun Solaris, HP e IBM por nombrar algunos. Estos proyectos de software libre y de código abierto son de hecho empresas que cuentan con una gran infraestructura especializada para cumplir con

el ciclo de vida del software y garantizar la funcionalidad del producto antes de ser liberado al mercado. Si bien, estos sistemas operativos pueden ser usados de acuerdo al esquema de licencia GPL, algunas funcionalidades especializadas de estos productos son limitados a su uso por un tiempo determinado hasta que el usuario o cliente se suscriba a un contrato de soporte técnico el cual puede oscilar entre los US\$ 320 a US\$ 1,200 al año por cada servidor en funcionamiento dependiendo también del nivel de servicio y funcionalidad requerida. [17]

Además de las razones descritas, al aplicar el filtro de pre selección de primera instancia que también se aplicó a las soluciones propietarias, causa que se descarte como alternativa a las soluciones de captura electrónica de datos con OMR basadas en licencia GPL para el procesamiento de evaluaciones de talento humano, ya que estas no demuestran aún ser maduras y confiables y no permiten proyectarse de manera comercial como un proveedor certificado y autorizado de soluciones y servicios por parte del fabricante con garantía de soporte formal.

3.3.2 SOLUCIONES PROPIETARIAS

Las soluciones propietarias analizadas muestran una gran madurez y vastas pruebas de confiabilidad y éxito a través de la exposición de casos de estudio en diferentes tamaños de proyectos especialmente de mediana y gran escala, tanto en escenarios para procesamiento a demanda como de misión crítica, implementados en diferentes países alrededor del mundo directamente por el fabricante de cada solución o a través de sus revendedores de valor agregado VAR⁴².

Siendo Kofax el fabricante que goza de mayor participación en el mercado de captura por lotes con una solución robusta, su costo para el procesamiento de formas fijas mediante OMR e ICR para específicamente evaluaciones de talento humano resulta ser el más alto por página entre las soluciones presentadas, y más aún, este costo se eleva considerablemente al pensar en Kofax Capture como la solución para proveer servicios de procesamiento y captura de datos con formas flexibles e inclusive formas fijas con una

⁴² VAR: Value Added Reseller.

mayor densidad de campos de selección por página, para lo cual Kofax sugiere el uso de módulos KTM especializados y licenciados por separado de manera muy específica de acuerdo a la necesidad o al contenido en la forma como tal, siendo esta una razón para descartar a Kofax Capture como opción para ser implementada.

IRIS cuenta con una gama interesante de productos en hardware para consumo lo cual lo hace una opción atractiva por la interoperabilidad existente entre sus mismos productos, sin embargo, las repetidas ocasiones que se solicitó información y una versión demostrativa de su solución a IRIS a través de su página de contacto en internet, no fue exitoso, razón por la cual no pudo este producto ser analizado con mayor detalle como una solución posible de ser implementada.

Según menciona la organización AIIM, IRIS está enfocada en desarrollar y atender mercados europeos y concentra sus esfuerzos de investigación y desarrollo hacia los idiomas de dichos países, lo que probablemente sus soluciones aún no están listas para incursionar en mercados latinos.

La gerente comercial de ABBYY Rusia a cargo de la región sudamericana, respondió de manera oportuna a la solicitud completada a través de su portal en internet, proveyendo con amplia información de producto y acceso a una versión demostrativa de ABBYY FlexiCapture. Posteriormente, durante una primera teleconferencia realizada, se indicó que en Ecuador no existía al momento un socio de negocio autorizado ni certificado de ABBYY, lo que generó amplia expectativa entre las partes para iniciar relaciones comerciales y brindó el apoyo necesario para poder participar en el entrenamiento de ABBYY FlexiCapture con la finalidad de obtener la certificación y calificación como VAR.

ABBYY FlexiCapture ofrece funcionalidades tecnológicas muy atractivas combinadas con un buen costo por página (*comparado con Kofax Capture solamente*), además de que la misma suite de producto incluye las herramientas necesarias para el diseño de formas fijas y plantillas flexibles las cuales son procesadas con la misma solución, siendo este un factor diferenciador respecto a las otras marcas analizadas que ofrecen productos complementarios especializados para el procesamiento de facturas como es el caso de IRISCapture Pro for Invoices, clasificación e indexación de documentos usando

IRISCapture Pro Sort & Index ó componentes adicionales para el procesamiento de documentos con características específicas como en el caso de Kofax Capture con Kofax Transformation Modules.

A pesar de que el objetivo de esta tesis está orientado al análisis de implementar EDC enfocándose como caso de estudio el procesamiento de evaluaciones de talento humano, es decir, formas fijas solamente, ABBYY FlexiCapture se presenta como la mejor alternativa con un horizonte más amplio para ofrecer servicios de captura electrónica de datos con diferentes tipos de documentos indistintamente del tipo de negocio, en donde no solo existen formas fijas sino también formatos flexibles inclusive con tablas de una o más páginas, lo cual hace que esta solución sea definitivamente la seleccionada para ser implementada.

Las dos ediciones de ABBYY FlexiCapture, Independiente y Distribuida, certifican responder a las más altas exigencias del mercado para cubrir las necesidades de captura electrónica de datos, con altos estándares de rendimiento, rentabilidad, facilidad de implementación y desarrollo, y aplicabilidad en proyectos tanto para la pequeña y mediana empresa hasta soluciones de gran escala en empresas geográficamente dispersas y gobierno.

Al momento, ABBYY Rusia tiene socios VAR en Argentina, Chile, Colombia, Paraguay, Uruguay y ahora en Ecuador, además de contar con presencia de marca como ABBYY en Brasil en donde efectivamente se llevó a cabo el entrenamiento y se obtuvo la certificación por parte de ABBYY Rusia.

4. Implementación de la Solución para Captura Electrónica de Datos

Habiendo seleccionado ABBYY FlexiCapture versión 10 como la solución a ser implementada para captura electrónica de datos, el presente capítulo describe las características y funcionalidades consideradas e implementadas exclusivamente para el procesamiento de evaluaciones de talento humano, generando mediciones de desempeño del sistema y comparativas respecto al proceso manual de captura de datos.

Conforme se van presentando las características primordiales de la solución, paralelamente se exponen los principios que fueron implementados para cumplir con las necesidades y requerimientos de procesamiento de un proyecto real de captura electrónica de datos para evaluaciones de talento humano.

4.1 Descripción del Proyecto

Por razones de confidencialidad al contrato de procesamiento de evaluaciones de talento humano con la organización cliente, cierta información como los nombres de los instrumentos de evaluación, las formas de respuesta, métodos de calificación y formatos de generación de resultados no son descritos explícitamente, razón por lo cual han sido sustituidos con información genérica en donde es requerido.

El proyecto de procesamiento del cual se tomó la información para el desarrollo del presente capítulo, operó bajo los siguientes requerimientos:

- Instrumentos de evaluación utilizados: 6.
- Cantidad de evaluados por instrumento de evaluación:
 - Forma 1: 1,034 evaluados.
 - Forma 2: 1,019 evaluados.
 - Forma 3: 77 evaluados.
 - Forma 4: 1,096 evaluados.
 - Forma 5: 77 evaluados.
 - Forma 6: 445 evaluados.
 - *Total: 3,748 evaluados.*

- Número de preguntas y opciones de respuesta por pregunta:
 - Forma 1: 40 preguntas con 5 opciones de respuesta cada una.
 - Forma 2: 71 preguntas con 5 opciones de respuesta cada una.
 - Forma 3: 86 preguntas con 5 opciones de respuesta cada una.
 - Forma 4: 105 preguntas con 5 opciones de respuesta cada una.
 - Forma 5: 30 preguntas con 5 opciones de respuesta cada una.
 - Forma 6: 30 preguntas con 5 opciones de respuesta cada una.
- Tipos de respuestas:
 - De selección múltiple de única respuesta.
- Métodos para la calificación automática de evaluaciones:
 - Batería de respuestas correctas.
 - Sumatoria nominal de las respuestas seleccionadas agrupadas por rangos de preguntas.
- Formatos de los entregables:
 - Base de datos en archivos de texto para cada instrumento de evaluación, el cual incluye identificador de evaluado con las selecciones individuales y nominales de respuesta a cada pregunta.
 - Base de datos en archivos de Microsoft® Excel para cada instrumento de evaluación, el cual incluye identificador de evaluado con la calificación respectiva de la evaluación.
 - Imágenes indexadas con identificador del evaluado, tipo de prueba y procedencia de la evaluación.
- Tiempo para generación de entregables: Hasta 72 horas según contrato.

4.2 Tipos de Instalación

La arquitectura moderna de ABBYY FlexiCapture 10 permite ser implementada para ofrecer una solución eficiente de captura electrónica de datos en un amplio rango y tipos de escenarios, de acuerdo al tipo de instalación elegida, para lo que propone dos alternativas de instalación con las que puede funcionar, cuya selección primaria se basa en el volumen potencial de documentos que se procesan por día, así como también la

cantidad de operadores que estarán involucrados en el procesamiento de documentos como tal.

4.2.1 INSTALACIÓN INDEPENDIENTE

Trabaja a manera de una aplicación de escritorio en donde todas las etapas del procesamiento suceden en una estación de trabajo. Está diseñada para atender las necesidades de captura electrónica de datos en proyectos pequeños o medianos del tipo departamental o en PYMEs⁴³, ofreciendo una instalación sencilla y un flujo de trabajo para el procesamiento documental automatizado con estadísticas sencillas de desempeño. Esta instalación incluye las siguientes estaciones:

Estación del Administrador

Permite la creación y configuración de plantillas y reglas documentales para el procesamiento y su administración. Puede también ser utilizada para capturar y procesar documentos.

Estación del Operador

Permite solamente la captura y el procesamiento en sí de los documentos sin que el operador tenga acceso a la administración y configuración de las plantillas documentales.

Un factor guía para elegir este tipo de instalación es que el volumen de documentos a ser procesados se encuentre entre 1,000 y 3,000 páginas por día y que el número de operadores de procesamiento involucrados globalmente sean máximo 3 usuarios.

4.2.2 INSTALACIÓN DISTRIBUIDA

Este tipo de instalación de ABBYY FlexiCapture 10 incluye un conjunto de componentes de servidor dedicados y distribuidos para el procesamiento de datos, así como estaciones de operación distribuidas (clientes) para los diferentes pasos del flujo de trabajo detallado más adelante y sobre los que operan los usuarios. Está diseñada para satisfacer las necesidades de procesamiento en proyectos de mediano y gran volumen tanto corporativo como gubernamental, de servicios BPO y de misión crítica.

⁴³ PYMEs: Pequeñas Y Medianas Empresas.

La arquitectura distribuida contempla características de alta escalabilidad con balanceo de cargas para asegurar una máxima productividad, funcionalidad de alta disponibilidad para prevenir fallas en el sistema, administración de usuarios y privilegios, operación remota, monitoreo, estadísticas y reportes de desempeño.

Un factor guía a considerar para elegir este tipo de instalación es que el volumen documental supere las 3,000 páginas por día involucrando a 3 o más operadores de procesamiento globalmente, por lo que este tipo de instalación es la seleccionada para ser implementada.

4.3 Arquitectura Distribuida

Los siguientes servidores y estaciones de operación brevemente descritas son las que están incluidas en la instalación distribuida de ABBYY FlexiCapture 10:

4.3.1 SERVIDOR DE APLICACIÓN

Encargado de ejecutar las tareas intensivas y demandantes de recursos como la importación de imágenes posterior a su digitalización, reconocimiento, clasificación, extracción de datos, aplicación y ejecución de reglas de validación y la exportación de datos e imágenes a destinos definidos. Su arquitectura es verdaderamente escalable distribuyendo las tareas de procesamiento a lo largo de todas las estaciones de procesamiento disponibles.

4.3.2 SERVIDOR DE PROCESAMIENTO

Se encarga de administrar y distribuir las colas de tareas a lo largo de las diferentes estaciones según su prioridad.

4.3.3 SERVIDOR DE LICENCIAMIENTO

Asegura que la operación de las estaciones y servidores estén de acuerdo al ámbito de licenciamiento además de contabilizar el volumen de páginas procesado.

4.3.4 SERVIDOR DE BASE DE DATOS

Mantiene y administra la información de los usuarios y sus privilegios, lotes documentales, operaciones de procesamiento, estadística y los datos extraídos como tal.

4.3.5 ESTACIÓN DE CONFIGURACIÓN DE PROYECTOS

Permite la creación, configuración y administración de proyectos para captura electrónica de datos, lo que incluye perfiles de importación o captura de imágenes, definiciones de plantillas documentales, configuración del flujo electrónico de trabajo durante el procesamiento, tipos de lotes, entre otros.

4.3.6 ESTACIÓN DE MONITOREO Y ADMINISTRACIÓN

Permite mediante un navegador de internet administrar bases de datos, asignar usuarios y sus privilegios a proyectos y tipos de lotes, generar reportes estadísticos a partir de las métricas de desempeño del sistema, su funcionamiento y costos de operación como tal.

4.3.7 ESTACIÓN DE DIGITALIZACIÓN

Permite la captura y separación de documentos desde diferentes orígenes, control de calidad y la edición de imágenes posterior a su captura. La captura de documentos puede realizarse de manera remota a través de internet directamente hacia FlexiCapture.

4.3.8 ESTACIÓN DE VERIFICACIÓN DE DATOS

Permite la verificación de datos a bajo nivel, es decir, la verificación de caracteres reconocidos inciertamente. Esta estación puede ser ejecutada a través de un navegador de internet, permitiendo que la verificación

4.3.9 ESTACIÓN DE VERIFICACIÓN

Ejecuta la verificación de datos en alto nivel, es decir, errores en reglas, excepciones en datos, confirmación de exportación y solicitud de digitalización fuera del flujo de proceso.

4.4 Descripción Funcional de la Solución

Los servidores de ABBYY FlexiCapture 10 pueden coexistir en un mismo computador (servidor) o ser asignados de manera individual a más de un computador para balancear las cargas de trabajo. Las estaciones de operación de ABBYY FlexiCapture 10 de manera similar pueden ser ejecutadas en un mismo computador o ser repartidas estratégicamente en varios computadores, considerando que el número máximo de estaciones de operación disponibles está determinado por el alcance de la licencia adquirida y que se consumen bajo el principio de concurrencia el cual contabiliza el número de estaciones que están en ejecución sin importar de cual computador en sí se está ejecutando la estación.

La configuración actualmente implementada ofrece un potencial de procesamiento de hasta 4,000 evaluaciones al día durante una jornada laboral, la cual incluye por licenciamiento 4 operadores de verificación (verificación de datos y verificación acomodados a demanda), 1 operador de digitalización, 1 estación de configuración de proyectos y un número virtualmente ilimitado de estaciones de procesamiento.

Para cumplir con los requerimientos del cliente en el actual proyecto de procesamiento de evaluaciones de talento humano, se requirió contar con únicamente 3 operadores en total, representando con la siguiente figura:

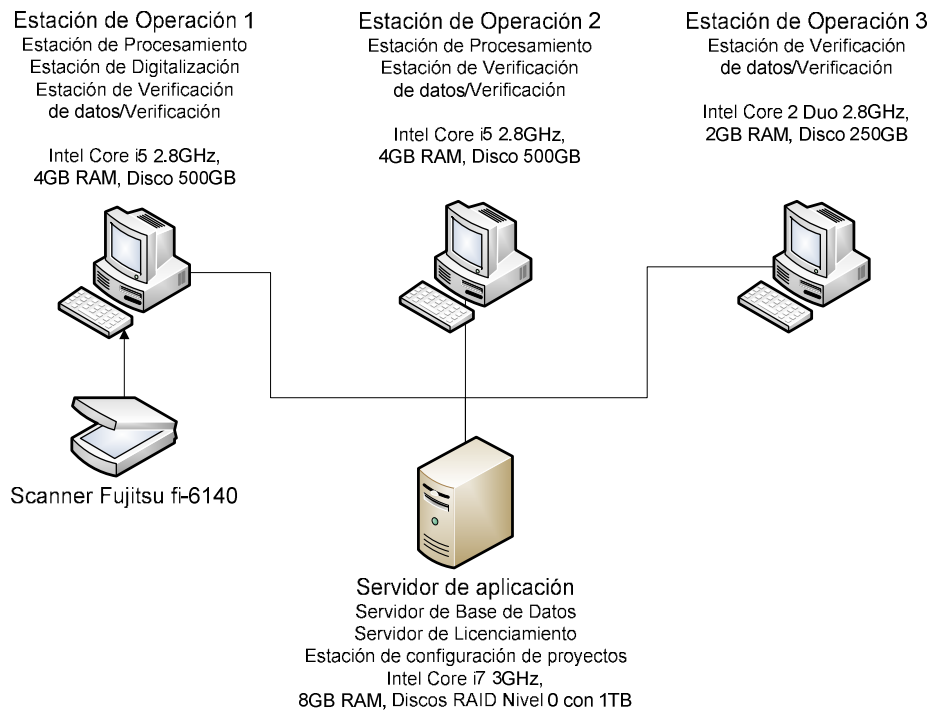


FIGURA 4-01: DIAGRAMA FUNCIONAL DE LA IMPLEMENTACIÓN DE ABBYY FLEXICAPTURE 10. [A]

4.5 Configuración y Flujo de Trabajo de ABBYY FlexiCapture 10

ABBYY FlexiCapture 10 comprende dos macro procesos en total para poder ofrecer el servicio de captura electrónica de datos, los que son claramente identificados como:

1. Configuración.
2. Procesamiento.

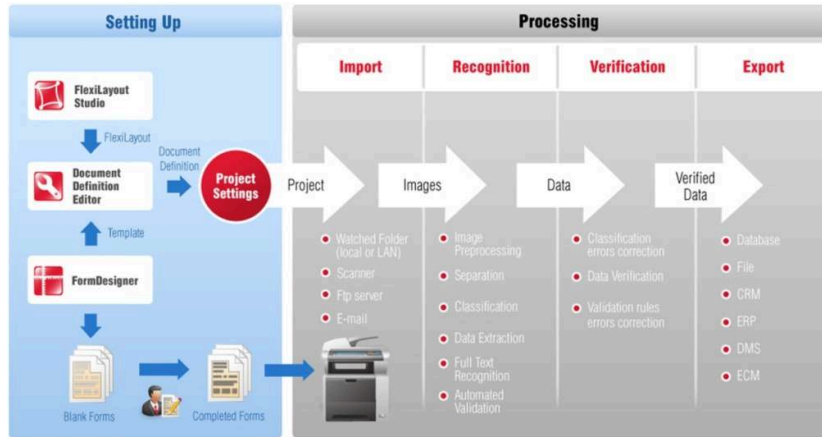


FIGURA 4-02: FLUJO DE TRABAJO DE ABBYY FLEXICAPTURE 10. [B]

4.5.1 CONFIGURACIÓN

Engloba todas las actividades relacionadas al diseño, configuración y puesta a punto de proyectos de procesamiento para ABBYY FlexiCapture 10.

Un proyecto se entiende como el conjunto de definiciones documentales los cuales guardan algún tipo de relación entre sí. Esta relación no necesariamente tiene que ver con el contenido en sí de los documentos, como en el caso en donde se estaría prestando servicios BPO para procesamiento de evaluaciones de talento humano a diferentes clientes, en donde para cada cliente se consideraría como un proyecto independiente. Para el caso en donde una empresa implementa internamente FlexiCapture, un proyecto podría contener diferentes definiciones documentales que en efecto guardan alguna relación entre ellos, sea departamental o por un proceso específico de negocio como por ejemplo, compras a proveedores o importaciones.

Una definición documental es el conjunto de parámetros y reglas de validación que identifican a una clase documental específica como por ejemplo una factura o una encuesta, y le permiten a FlexiCapture durante el procesamiento encontrar la localización relativa o absoluta de los campos o elementos de interés cuyos datos deben ser extraídos.

Las tareas comúnmente realizadas para la configuración de proyectos se basan en la estación de configuración de proyectos y herramientas adicionales incluidas en ABBYY FlexiCapture, siendo estas:

- ABBYY FormDesigner.
 - Diseño de formas fijas vacías.
- ABBYY FlexiLayout Studio.
 - Creación de plantillas documentales flexibles.
- Document Definition Editor.
 - Configuración de definiciones documentales fijas y flexibles.
 - Publicación de Proyectos.
- Estación de configuración de proyectos.
 - Definición de tipos de lotes.
 - Parámetros para importación de imágenes.
 - Parámetros para exportación de datos e imágenes.
 - Configuración del flujo electrónico de trabajo.

Aplicación al Proyecto

Con ABBYY FormDesigner se diseñaron las 6 formas con segregación de color gris para los instrumentos de evaluación, las que una vez aprobadas fueron impresas en las cantidades requeridas según el número de evaluados.

Cada una de las formas contiene las marcas de anclaje requeridas para una forma fija además de:

- Cabecera de forma:
 - Identificador textual de instrumento de evaluación.
 - Ejemplo del mapa de caracteres con instrucciones para el llenado.
 - Campo para nombres con espacio hasta para 33 caracteres.
 - Campo para apellidos con espacio hasta para 33 caracteres.
 - Campo para fecha de evaluación en formato “dd-mm-aaaa”.
 - Campo para número de cédula con espacio para 10 dígitos.
- Sección para la selección de respuestas a preguntas:
 - Preguntas numeradas con el número de opciones requerido según el tipo de instrumento.
- Pié de forma:
 - Identificador numérico de forma.

- Importación de documentos.
- Reconocimiento:
 - Clasificación y ensamblaje de documentos.
 - Extracción de datos.
- Verificación:
 - Verificación de datos (bajo nivel).
 - Verificación (alto nivel).
- Exportación de datos e imágenes.

Importación de documentos

Usando la estación de digitalización de FlexiCapture, los documentos a ser procesados deben ser importados utilizando alguno de los siguientes orígenes:

- Un dispositivo de escaneo multifuncional (MFP) ó un escáner de escritorio ó de producción que soporta protocolos de digitalización WIA, TWAIN ó ISIS.
- Importación pasiva de imágenes desde una o más carpetas desde un disco duro o una ubicación de red.
- Importación activa de imágenes usando perfiles de importación.
- Importación de documentos adjuntos en correo electrónico de Microsoft® Exchange o servidores de email POP3⁴⁴.
- Importación de archivos desde un servidor FTP.

Previa a la importación, los documentos son organizados en grupos lógicos conocidos como lotes documentales, para los cuales y opcionalmente se definen variables adicionales llamadas parámetros de registro, que son completados de manera manual ó automática respecto a cada lote tales como el nombre de usuario de digitalización, la estación de la cual se realizó la importación ó información propia del lote que servirán para garantizar la trazabilidad de cada documento a lo largo de todo el procesamiento.

⁴⁴ POP3: Post Office Protocol 3.

A la vez que las imágenes están siendo capturadas en la estación de digitalización, visualmente una por una son controladas en calidad y opcionalmente ejecutadas tareas para procesamiento de imágenes, tales como:

- Separación automática de documentos basada en páginas en blanco, códigos de barras ó número fijo de páginas.
- Destramado, enderezado y eliminación de fondo de imagen.
- Rotación de imagen.
- Conversión de imágenes en color y escala de gris a imágenes blanco y negro o en negativo.
- Redacción u ocultamiento de áreas sensibles.

Finalmente, las imágenes listas para ser procesadas son enviadas a un proyecto previamente seleccionado ó predefinido hacia el servidor de aplicación de ABBYY FlexiCapture 10.

Aplicación al Proyecto

Dada la naturaleza de las evaluaciones que comprenden una sola página por evaluación y tipo de instrumento, se diseñó un tipo de lote que no realice separación de documentos ni clasificación, fijando 3 parámetros de registro para cada lote recibido, correspondientes a la ciudad, el aula y turno en donde se llevaron a cabo las evaluaciones.

Las evaluaciones completadas fueron digitalizadas por un operador de digitalización usando un escáner de escritorio marca Fujitsu modelo fi-6140 que cuenta con una capacidad de producción diaria de 4,000 hojas, alimentador automático de documentos para 50 hojas y una velocidad de digitalización de 40 imágenes por minuto en modo simplex ó de 80 imágenes por minuto en modo dúplex a 300 DPI de resolución, la cual es la resolución óptima para captura electrónica de datos.

Si bien, potencialmente se digitalizan 40 imágenes por minuto y el alimentador de páginas del escáner usado está diseñado para contener 50 páginas, resulta que mientras estas siguen pasando a través del escáner, se puede seguir añadiendo las páginas restantes del mismo lote al alimentador si es que fuera mayor a 50 páginas,

con lo que el tiempo de digitalización como tal no es el crítico, sino más bien la preparación de los lotes previos a ser digitalizados, lo que puede exigir quitar grapas, clips u hojas adhesivas que podrían contener, además que por control de calidad se cuentan físicamente las páginas para validar con el conteo reportado en la estación de digitalización.

A pesar del tiempo requerido que se consume previo a la digitalización, este proceso trabaja en paralelo respecto al de los operadores de verificación de datos, lo que quiere decir que dichos operadores podrán iniciar a trabajar cuando al menos un lote digitalizado haya sido enviado al servidor de procesamiento y se haya reconocido.

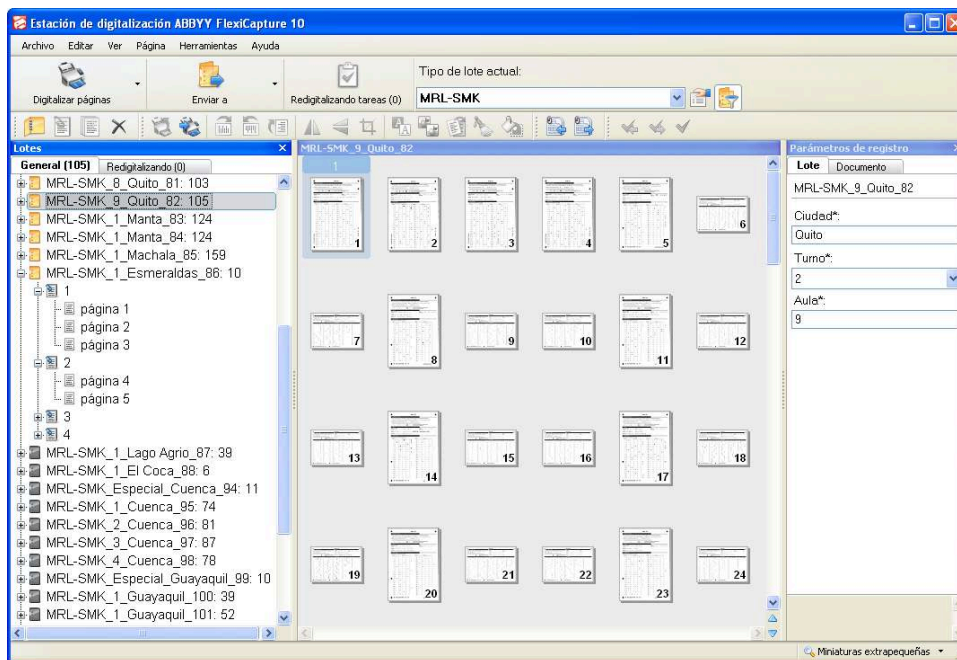


FIGURA 4-07: ESTACIÓN DE DIGITALIZACIÓN DE ABBYY FLEXICAPTURE 10. [A]

Clasificación y Ensamblaje de Documentos

FlexiCapture recurre al análisis de cada imagen importada para determinar cuál de las definiciones documentales coincide con los elementos de identificación encontrados en el documento analizado para proceder con la extracción de datos, además de verificar que las páginas existentes en un mismo documento estén en el orden esperado caso contrario serán marcados con error de ensamblaje.

En esta etapa opcionalmente se pueden usar árboles de clasificación documental al tratar con una cantidad considerable de clases y tipos documentales, con la finalidad de acelerar procesos de clasificación e identificación de documentos de manera preliminar para agruparlos de manera lógica y multinivel, por ejemplo, un primer nivel separa contratos y facturas, un segundo nivel separa facturas por proveedores y un tercer nivel separa por monto facturado.

Extracción de Datos

Esta etapa comprende la aplicación en sí de las tecnologías de reconocimiento de caracteres OCR e ICR, marcas con OMR y códigos de barras con BCR a cada uno de los campos identificados en el documento y la asociación de cada uno a un campo electrónico en el formulario de datos para su posterior verificación y ejecución de reglas de validación.

Hasta esta etapa el flujo de trabajo de FlexiCapture es totalmente desatendida o automáticas y es crítica en cuanto a la precisión de reconocimiento de datos, ya que al trabajar con una imagen óptima, la calidad del OCR puede representar una precisión del 99% para alfabetos latinos, un 95% para el ICR y 100% para OMR cuando el formulario está optimizado para la lectura por computador y ha sido completado de manera correcta.

Aplicación al Proyecto

Al momento que el servidor de procesamiento recibe las imágenes enviadas desde la estación de digitalización, son repartidas a todas las estaciones de procesamiento disponibles para arrancar con la fase de reconocimiento en sí, identificando a cada uno de los documentos con el nombre de la definición documental con la que encontró correspondencia entre las existentes en el proyecto. Al finalizar se pueden tener métricas iniciales respecto al nivel de confianza de reconocimiento, el porcentaje y números de caracteres inciertos y de caracteres para ser verificados por el operador de verificación.

ANÁLISIS COSTO/BENEFICIO DE IMPLEMENTAR TECNOLOGÍAS DE CAPTURA ELECTRÓNICA DE DATOS (EDC). ÁREA DE ESTUDIO: EVALUACIONES AL TALENTO HUMANO

#	Nombre	Siguiente etapa	Nivel de confianza	Caracteres inciertos	Caracteres para la verificación	Identificador de documentos	Notas de procesamiento
1	Forma 1	Verificación	86% (285 de 332)	10% (33 de 332)	14% (47 de 332)	00000001	
2	Forma 2	Verificación	96% (689 de 721)	3% (25 de 721)	4% (32 de 721)	00000002	
3	Forma 4	Verificación	95% (685 de 722)	4% (31 de 722)	5% (37 de 722)	00000003	
4	Forma 6	Verificación	92% (323 de 350)	3% (12 de 350)	8% (27 de 350)	00000004	
5	Forma 1	Verificación	91% (299 de 327)	4% (12 de 327)	9% (28 de 327)	00000005	
6	Forma 2	Verificación	96% (693 de 722)	3% (22 de 722)	4% (29 de 722)	00000006	
7	Forma 4	Verificación	96% (693 de 723)	3% (24 de 723)	4% (20 de 723)	00000007	
8	Forma 3	Verificación	96% (691 de 721)	3% (25 de 721)	4% (30 de 721)	00000008	
9	Forma 4	Verificación	95% (688 de 722)	4% (29 de 722)	5% (34 de 722)	00000009	
10	Forma 5	Verificación	90% (296 de 329)	5% (17 de 329)	10% (33 de 329)	00000010	
11	Forma 1	Verificación	93% (300 de 324)	2% (7 de 324)	7% (24 de 324)	00000011	
12	Forma 2	Verificación	95% (687 de 720)	4% (27 de 720)	5% (33 de 720)	00000012	
13	Forma 4	Verificación	96% (702 de 734)	3% (25 de 734)	4% (32 de 734)	00000013	
14	Forma 1	Verificación	91% (296 de 326)	4% (13 de 326)	9% (30 de 326)	00000014	
15	Forma 2	Verificación	96% (689 de 721)	3% (24 de 721)	4% (32 de 721)	00000015	
16	Forma 4	Verificación	96% (691 de 722)	3% (23 de 722)	4% (31 de 722)	00000016	
17	Forma 1	Verificación	93% (306 de 329)	2% (8 de 329)	7% (23 de 329)	00000017	
18	Forma 2	Verificación	96% (696 de 724)	3% (21 de 724)	4% (28 de 724)	00000018	
19	Forma 4	Verificación	97% (702 de 725)	2% (16 de 725)	3% (23 de 725)	00000019	
20	Forma 6	Verificación	94% (341 de 361)	1% (2 de 361)	6% (20 de 361)	00000020	
21	Forma 1	Verificación	91% (299 de 327)	3% (11 de 327)	9% (28 de 327)	00000021	
22	Forma 2	Verificación	96% (693 de 722)	3% (22 de 722)	4% (29 de 722)	00000022	
23	Forma 4	Verificación	97% (687 de 711)	3% (20 de 711)	3% (24 de 711)	00000023	

FIGURA 4-08: DOCUMENTOS EN UN LOTE LISTO PARA VERIFICACIÓN. [A]

Este proceso también funciona en paralelo de manera similar a como sucede con la estación de digitalización, de tal manera que los operadores de verificación de datos pueden iniciar a trabajar cuando al menos un lote de documentos ha sido ya reconocido, mientras que el resto de lotes siguen procesándose y entrando en cola a la espera de ser verificados.

El siguiente cuadro presenta el tiempo aproximado de reconocimiento usando la actual infraestructura implementada.

Tiempo de Reconocimiento		
Descripción	Valor	Unidad
Número de estaciones de reconocimiento	3	estaciones
Total de evaluaciones a reconocer	3,748	evaluaciones
Número de evaluaciones reconocidas por segundo	3	evaluaciones
Tiempo requerido para reconocimiento (1)	1,249.33	segundos
Tiempo potencial utilizado para reconocimiento	20.82	minutos

No fue requerida la aplicación de un árbol de clasificación documental ya que el número de definiciones documentales es limitado tan solo a 6. Adicionalmente, todos los documentos se conforman de una sola página por lo que no es necesario ejecutar tareas de ensamblaje de documentos.

Verificación de Datos (bajo nivel)

Usando la estación de verificación de datos, los datos extraídos que no superan el nivel de confianza configurado, son enviados a verificación grupal en primera

instancia, mostrando todos aquellos caracteres que individualmente son idénticos entre sí para todos los documentos de un mismo lote, tales como marcas de selección y números dígitos. En esta etapa el operador hace las correcciones sobre aquellos caracteres que visualmente no coinciden con el valor de grupo reconocido en el procesamiento.

A continuación se ejecuta la verificación de campos individuales que contienen caracteres dudosos los cuales son mostrados en color rojo por no haber superado el nivel de confianza establecido en la definición documental luego de haber sido reconocidos. Adicional y automáticamente las advertencias y errores provenientes de las reglas de validación en el contexto del campo son mostradas en caso de no cumplirse, garantizando así la exactitud, confiabilidad y precisión de los datos reconocidos por verificar para minimizar el error humano.

El operador de verificación de datos en caso de no poder dar solución a un mensaje de advertencia o error generado por las reglas de validación, está en capacidad de posponer la verificación de dicho campo causando que el documento se marque y sea presentado al operador de la siguiente etapa de verificación de alto nivel para obtener solución.

Las reglas de validación que pueden aplicarse a los datos son de los siguientes tipos:

- Reglas no programáticas:
 - Admisión de selección múltiple y/ó vacía a campos de selección.
 - Admisión de campos en blanco.
 - Validación de datos respecto al tipo de dato esperado.
 - Longitud máxima de campo.
 - Lista o base de datos con valores permitidos.
 - Formato basado en expresión regular.
 - Validación por autocorrección, reemplazo de caracteres y texto.
 - Validación de comprobación mediante búsqueda exacta, búsqueda difusa ó importación de valores entre los campos capturados y una base de datos.

- Verificación de campos calculados.
- Comparación de campos idénticos.
- Codificación y reemplazo de valores desde una lista predeterminada.
- Reglas programáticas con scripts en lenguajes .Net, JScript® y VBScript.
 - Cualquier regla de validación sobre un campo individual o entre varios campos.

Aplicación al Proyecto

Los operadores de verificación ejecutaron la verificación grupal de marcas de selección y números dígitos así como la verificación en campo de caracteres dudosos para nombres y apellidos.



FIGURA 4-09: VERIFICACIÓN DE DATOS EN GRUPO PARA MARCAS DE SELECCIÓN. [A]

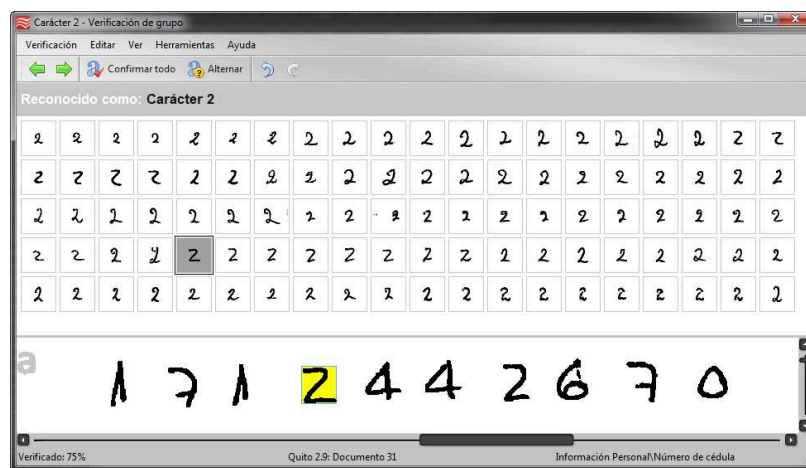


FIGURA 4-10: VERIFICACIÓN DE DATOS EN GRUPO PARA EL NÚMERO DÍGITO 2. [A]

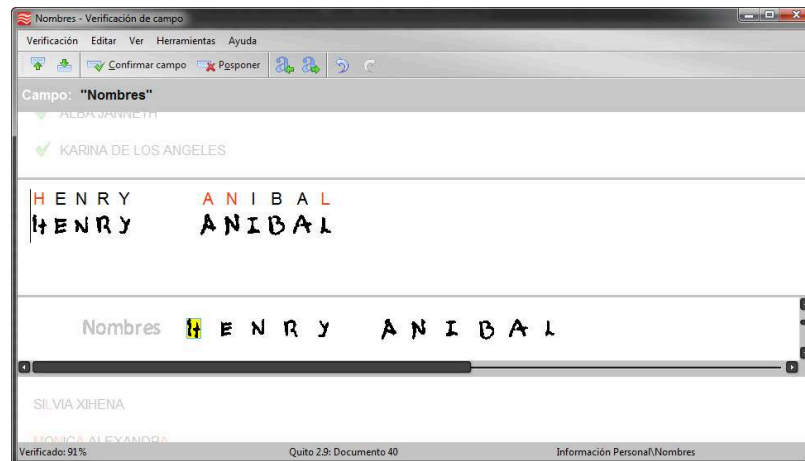


FIGURA 4-11: VERIFICACIÓN DE DATOS EN CAMPO SIN ERROR DE VALIDACIÓN. [A]

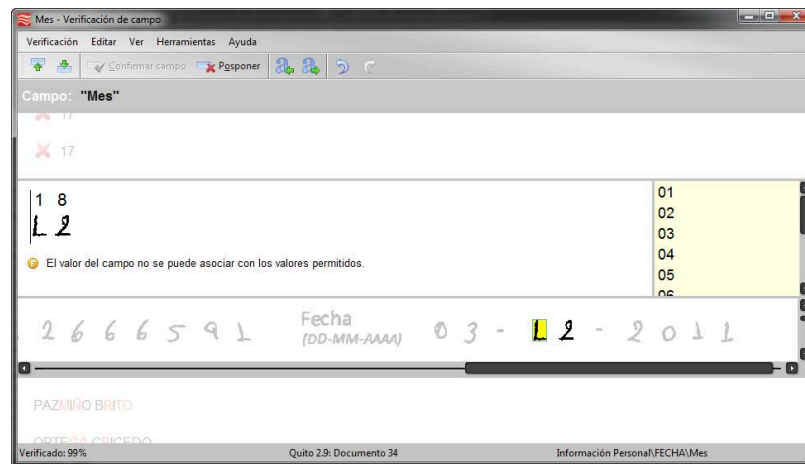


FIGURA 4-12: VERIFICACIÓN DE DATOS EN CAMPO CON ERROR DE VALIDACIÓN EN VALOR. [A]

Las reglas implementadas en cada definición documental en la fase de configuración del proyecto y que corrieron durante la verificación de datos fueron:

- Validación de formato de fecha.
- Validación de fecha válida entre los rangos esperados.

Verificación (alto nivel)

Es una etapa dentro del flujo electrónico de documentos de FlexiCapture que permite examinar de manera manual la exactitud del proceso de reconocimiento sobre los campos y datos comparando los campos situados sobre la imagen contra los campos en el formulario electrónico. Esta etapa es opcional y puede ser activada

para todos los campos o algunos seleccionados según la criticidad de la información, siendo esta la última en donde el operador de verificación corrige los errores marcados sobre los documentos y campos individuales para así garantizar total precisión en los datos que serán exportados.

Un operador de verificación senior es quien puede ver a través de la estación de verificación todos los documentos pertenecientes a un lote, y examinar la causa por la que han sido marcados con uno de los siguientes íconos visuales:

- Documento con bandera de color amarillo como señal de advertencia.
- Documento con banderas de color rojo como señal de error.
- Documento con una “x” dentro de un círculo rojo como error de ensamblaje.

#	Nombre	Siguiente etapa	Nivel de confianza	Caracteres incorrectos	Caracteres para la verificación	Identificador de documentos	Notas de procesamiento
1	Forma 1	Exportar	100% (330 de 330)	9% (31 de 330)	0% (0 de 330)	00000001	
2	Forma 2	Exportar	100% (719 de 719)	3% (23 de 719)	0% (0 de 719)	00000002	
3	Forma 4	Exportar	100% (720 de 720)	4% (29 de 720)	0% (0 de 720)	00000003	
4	Forma 6	Exportar	100% (348 de 348)	3% (10 de 348)	0% (0 de 348)	00000004	
5	Forma 1	Verificación	99% (326 de 327)	3% (11 de 327)	1% (1 de 327)	00000005	
6	Forma 2	Verificación	99% (721 de 722)	3% (19 de 722)	1% (1 de 722)	00000006	
7	Forma 4	Exportar	100% (721 de 721)	3% (22 de 721)	0% (0 de 721)	00000007	
8	Forma 3	Exportar	100% (719 de 719)	3% (21 de 719)	0% (0 de 719)	00000008	
9	Forma 4	Exportar	100% (720 de 720)	3% (25 de 720)	0% (0 de 720)	00000009	
10	Forma 5	Exportar	100% (327 de 327)	5% (15 de 327)	0% (0 de 327)	00000010	
11	Forma 1	Exportar	100% (322 de 322)	2% (5 de 322)	0% (0 de 322)	00000011	
12	Forma 2	Exportar	100% (718 de 718)	3% (23 de 718)	0% (0 de 718)	00000012	
13	Forma 4	Exportar	100% (732 de 732)	3% (23 de 732)	0% (0 de 732)	00000013	
14	Forma 1	Exportar	100% (324 de 324)	3% (11 de 324)	0% (0 de 324)	00000014	
15	Forma 2	Exportar	100% (719 de 719)	3% (22 de 719)	0% (0 de 719)	00000015	
16	Forma 4	Exportar	100% (720 de 720)	3% (21 de 720)	0% (0 de 720)	00000016	
17	Forma 1	Exportar	100% (327 de 327)	2% (7 de 327)	0% (0 de 327)	00000017	
18	Forma 2	Exportar	100% (722 de 722)	2% (17 de 722)	0% (0 de 722)	00000018	
19	Forma 4	Exportar	100% (723 de 723)	2% (12 de 723)	0% (0 de 723)	00000019	
20	Forma 6	Exportar	100% (359 de 359)	1% (2 de 359)	0% (0 de 359)	00000020	
21	Forma 1	Exportar	100% (325 de 325)	3% (9 de 325)	0% (0 de 325)	00000021	
22	Forma 2	Exportar	100% (720 de 720)	3% (20 de 720)	0% (0 de 720)	00000022	
23	Forma 4	Verificación	100% (714 de 714)	3% (18 de 714)	0% (0 de 714)	00000023	

FIGURA 4-13: DOCUMENTOS CON ADVERTENCIAS Y ERRORES DE VALIDACIÓN EN LA ESTACIÓN DE VERIFICACIÓN. [A]

Aplicación al Proyecto

De los 3 operadores, uno cuenta con los privilegios de verificador senior, quien está a cargo de ejecutar esta última instancia de verificación, confirmar que los lotes documentales están listos para ser exportados y ejecutar como tal la exportación certificando la inexistencia de error alguno.

En esta etapa del proceso se ejecutan reglas adicionales de validación no presentadas en la verificación de datos las cuales deben ser solucionadas documento por documento para aquellos que no se muestran como listos para ser exportados, tal como se presenta en la figura siguiente:

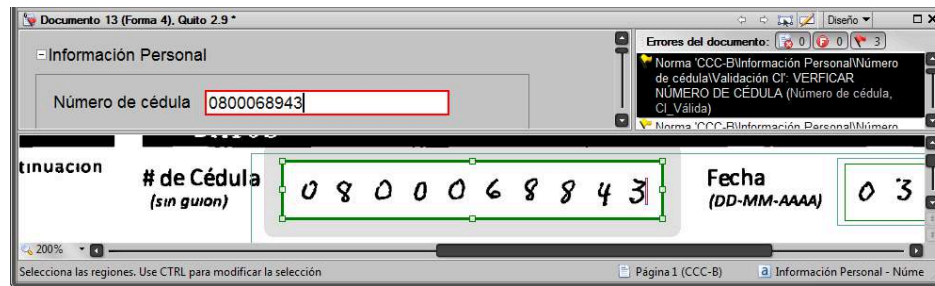


FIGURA 4-14: ERROR DE VALIDACIÓN PARA EL DÍGITO VERIFICADOR EN LA CÉDULA DE IDENTIDAD. [A]

Las reglas de validación que se implementaron para la presente etapa fueron las siguientes:

- Calificación automática de evaluaciones de acuerdo a los dos métodos requeridos por el cliente.
- Validación programática mediante script para el dígito verificador en la cédula de identidad.
- Búsqueda exacta de cédula de identidad en base de datos de evaluados.
- Advertencia al detectar selección vacía ó múltiple en una pregunta.

Exportación de Datos

Esta etapa de ejecución automática cuando todos los documentos están marcados listos para exportación, es la encargada de enviar los datos e imágenes a los diferentes destinos definidos durante la configuración de cada definición documental, pudiendo ser:

- Datos hacia bases de datos transaccionales, hojas de Excel, archivos de texto, archivos XML, archivos planos CSV ó archivos DBF.
- Imágenes indexadas de página simple o multi página con múltiples formatos como PDF, PDF/A, TIFF, BMP, JPG u otros hacia una estructura de archivos o repositorios documentales.

Aplicación al Proyecto

Los entregables requeridos por el cliente son creados automáticamente en esta etapa. La exportación a dos formatos de bases de datos e imágenes indexadas por número de cédula y clasificadas según los parámetros de registro de cada lote se la

inició a ejecutar a las 48 horas de haber recibido las evaluaciones para ser procesadas, lo que permitió cumplir con el tiempo de entrega de acuerdo a lo establecido en el contrato.

El conjunto de entregables finalmente son cargados a un servidor FTP del cual el cliente procede a descargarlos para proceder con el análisis psicométrico respectivo.

4.6 Retorno Sobre la Inversión de TI (IT ROI)

Siendo el principal objetivo para abordar un proyecto tecnológico, el lograr eficiencia operacional en procesos de negocio, el retorno sobre la inversión de TI es un parámetro de medición expresado porcentualmente, usado con gran frecuencia para cuantificar el impacto y beneficio monetario que logra una inversión de capital en el tiempo respecto a un proyecto de tecnología de la información.

El período de tiempo generalmente aceptado para la aplicación del ROI⁴⁵ de TI es de tres años, ya que se considera que la tecnología es prácticamente obsoleta luego de este período. Esta medición no evalúa los factores intangibles o cualitativos del impacto los cuales son los más difíciles de cuantificar, a pesar que un proyecto de tecnología de la información adecuado genera beneficios intangibles muy palpables.

El beneficio en sí no se encuentra solo en la captura de datos por sí sola. Hay muchos beneficios intrínsecos asociados a la captura electrónica de datos como la capacidad de indexar datos e imágenes en repositorios de almacenamiento electrónico que permiten la consulta eficiente de los mismos, la toma de decisiones oportuna con el apoyo de herramientas para automatizar procesos de negocio, la replicación y respaldo de datos y la oportunidad que tienen los usuarios para consumir la información.

Específicamente a continuación se desarrolla el cálculo del ROI para captura electrónica de datos usando ABBYY FlexiCapture 10 a partir de los indicadores de desempeño medidos en el proyecto de procesamiento descrito en este capítulo proyectado a un año

⁴⁵ ROI: Return On Investment.

para equiparar con el cálculo de costos de procesamiento manual presentado en el capítulo 2 con el título “El Costo del Procesamiento Manual a Mayor Escala”, e igualmente no se toman en cuenta otros factores de costos como espacio de operación, mobiliario y equipos de cómputo los cuales en ambos escenarios son existentes. Con este análisis se podrá comprobar el beneficio que existe al invertir en la implementación de ABBYY FlexiCapture para el procesamiento de evaluaciones de talento humano frente a la captura manual de datos.

4.6.1 FACTORES CONSIDERADOS PARA EL CÁLCULO DEL ROI

Volumen de páginas a ser procesadas

La instalación sugerida de ABBYY FlexiCapture para procesar un volumen anual de 120,000 formas fijas de una página cada una es del tipo independiente, y siendo conocido el hecho que un operador de verificación puede procesar hasta 1,000 páginas al día, la producción diaria que no supera las 500 páginas en este escenario requeriría contar con solo un operador de verificación (verificación de datos y verificación) más uno para digitalización.

ITEM		Resumen del procesamiento manual	
	Número de días laborables en 1 año		254
	Páginas a procesar al año		120,000
	Número de páginas a procesar por día		473
	Tiempo promedio en segundos para procesar 1 evaluación		420
	Horas día requeridas para el ingreso manual de datos		55.2
	Días laborables al año (con 10 días feriados)		254
	Operadores para el ingreso de datos requeridos por turno		7
	Salario del operador de ingreso de datos por año	\$	4,260.00
O	Costo anual de operadores para procesamiento manual	\$	29,820.00

FIGURA 4-15: RESUMEN DE PROCESAMIENTO MANUAL PLANTEADO EN EL CAPÍTULO 2. [A]

Implementación de la Solución

Para la implementación de ABBYY FlexiCapture es necesario contar con una licencia de software que anualmente pueda cubrir la cantidad de páginas a ser procesadas más un excedente de protección de 10,000 páginas, es decir, 130,000 páginas anuales. Adicionalmente se requiere de servicios profesionales para la implementación, entrenamiento y la compra de un escáner que en el escenario de procesamiento manual no existe.

Los computadores se asume que serían los mismos que se destinarían para el procesamiento manual, razón por la que tampoco están considerados en el cálculo.

ITEM Inversión en licenciamiento de software				
Inversión Inicial				
Software ABBYY FlexiCapture				
A1	ABBY FlexiCapture 10 para formas fijas, instalación independiente, 130,000 páginas al año autorenovables.			\$ 2,128.00
A2	Estación de operación adicional	1	\$ 710.00	\$ 710.00
A	Costo Total de Licenciamiento de Software			\$ 2,838.00
Hardware				
B1	Escaner con ADF para 2,000 páginas al día	1	\$ 1,495.00	\$ 1,495.00
B	Costo Total de Hardware			\$ 1,495.00
Servicios Profesionales				
C1	Implementación de la solución (<i>Calculado de A</i>)	1	25%	\$ 709.50
C2	Diseño de formas y definiciones documentales	4	\$ 250.00	\$ 1,000.00
C	Costo Total de Servicios Profesionales			\$ 1,709.50
Entrenamiento				
D1	Entrenamiento técnico para 2 personas	1	\$ 1,295.00	\$ 1,295.00
D	Costo Total de Entrenamiento			\$ 1,295.00
	Total Inversión Inicial			\$ 7,337.50

FIGURA 4-16: INVERSIÓN EN IMPLEMENTACIÓN DE SOFTWARE. [A]

Costo Anual de Operación

Dado que el cálculo del ROI se proyecta a 3 años como se explicó, la implementación de software requiere de un pago anual por concepto de soporte y mantenimiento de la solución el cual le otorga al cliente el derecho de acceder a las versiones más recientes del producto así como a solicitar soporte técnico cuando sea necesario. Parte del costo anual de operación es el costo de personal de operación respectivo para trabajar sobre la solución.

ITEM Costo anual de operación				
Soporte y Mantenimiento Anual				
E1	SMA Licencias de software (<i>Calculado de A</i>)	1	20%	\$ 567.60
E2	Hardware (<i>Calculado de B. Aplica a partir del año 2</i>)	1	20%	\$ 299.00
E	Costo Total de Soporte y Mantenimiento			\$ 866.60
Personal de Operación				
F1	Operadores	2	\$ 4,260.00	\$ 8,520.00
F	Costo Total de Personal			\$ 8,520.00
	Costo Total Anual			\$ 9,386.60

FIGURA 4-17: COSTOS ANUALES DE OPERACIÓN CON EDC. [A]

4.6.2 CÁLCULO DEL ROI

Los factores considerados son suficientes para calcular el retorno sobre la inversión realizada al implementar ABBYY FlexiCapture para la captura electrónica de evaluaciones de talento humano.

ITEM		Cálculo del ROI
Procesamiento Manual		
CPM	Costo anual de operadores para procesamiento manual	\$ 29,820.00
Sistema ABBYY FlexiCapture		
A	Costo Total de Licenciamiento de Software	\$ 2,838.00
B	Costo Total de Hardware	\$ 1,495.00
C	Costo Total de Servicios Profesionales	\$ 1,709.50
D	Costo Total de Entrenamiento	\$ 1,295.00
T1	Total Inversión Inicial	\$ 7,337.50
Costos Anuales		
E1	Costo Total de Soporte y Mantenimiento de Software	\$ 567.60
E2	Costo Total de Mantenimiento de Hardware (A partir del año 2)	\$ 299.00
F	Costo Total de Personal	\$ 8,520.00
T2	Costo Total Anual	\$ 9,386.60
Resumen de Costos en 3 Años		
C1	Año 1 (T1+T2-E2)	\$ 16,425.10
C2	Año 2 (T2)	\$ 9,386.60
C3	Año 3 (T2)	\$ 9,386.60
Resumen de Ahorros en 3 Años		
S1	Año 1 (CPM-C1)	\$ 13,394.90
S2	Año 2 (CPM-C2)	\$ 20,433.40
S3	Año 3 (CPM-C3)	\$ 20,433.40
Cálculo del ROI		
R1	Inversión Total en la Solución de EDC (C1+C2+C3)	\$ 35,198.30
R2	Total de Ahorros Generados (S1+S2+S3)	\$ 54,261.70
R3	Retorno Sobre la Inversión a 3 Años (R2/R1)	154%
R4	Meses para Recuperar la Inversión (12/R3)	7.8

FIGURA 4-18: ROI PARA LA SOLUCIÓN EDC CON ABBYY FLEXICAPTURE 10. [A]

La recuperación de la inversión menor a un año es indudablemente un resultado muy alentador y satisfactorio que justifica y alienta la adopción de ABBYY FlexiCapture para la captura electrónica de datos aplicadas a las evaluaciones de talento humano.

Al comparar el costo asociado al procesamiento manual de cada página respecto al costo de procesar cada página con ABBYY FlexiCapture como solución EDC, se genera un ahorro superior al 52% para el primer año y sobre el 69% a partir del segundo año, tal como se muestran en las siguientes figuras:

Costo por página (sin EDC)	
<i>En cualquier año</i>	
Páginas a procesar en el año	120,000
Operadores (7 recursos)	\$ 29,820.00
Costo por página al año (sin EDC)	\$ 0.2485

FIGURA 4-19: COSTO POR PÁGINA PROCESADA SIN EDC. [A]

Costo por página (con EDC)	
<i>En el primer año</i>	
Páginas a procesar en el año	130,000
Costo de software y servicios (A+C+D+E1)	\$ 6,410.10
Hardware (Depreciado a 3 años: B/3)	\$ 498.33
Operadores (2 recursos) (F1)	\$ 8,520.00
Costo por página en el primer año	\$ 0.1187
<i>A partir del año 2</i>	
Páginas a procesar en el año	130,000
Costo de software y servicios (E)	\$ 866.60
Hardware (Depreciado a 3 años: B/3)	\$ 498.33
Operadores (2 recursos) (F1)	\$ 8,520.00
Costo por página a partir del año 2	\$ 0.0760

FIGURA 4-20: COSTOS POR PÁGINA PROCESADA CON EDC. [A]

Ahorro generado en el costo por página	
<i>En el primer año</i>	
Costo por página procesada con EDC	\$ 0.1187
Costo por página procesada sin EDC	\$ 0.2485
Ahorro por página durante el primer año	-52.2%
<i>A partir del año 2</i>	
Costo por página procesada con EDC	\$ 0.0760
Costo por página procesada sin EDC	\$ 0.2485
Ahorro por página a partir del año 2	-69.4%

FIGURA 4-21: AHORROS GENERADOS RESPECTO AL COSTO POR PÁGINA PROCESADA. [A]

En escenarios de mayor volumen documental, el retorno sobre la inversión seguirá aún siendo atractivo debido a que el costo de licenciamiento por página se reduce conforme el volumen potencial de procesamiento se incrementa, obteniendo costos por página procesada más bajos al contar con la cantidad de operadores adecuada.

5. Conclusiones y Recomendaciones

5.1 Conclusiones

5.1.1 PRODUCTIVIDAD Y ESCALABILIDAD

El procesamiento manual para captura de datos demuestra ser costoso y difícilmente escalable y manejable, mientras que al ser necesario escalar con la solución implementada, hasta cierto punto no es necesario aumentar personal, como en el caso analizado en donde el procesamiento de 473 páginas al día, es prácticamente la productividad de un operador al 50%. Este factor permite generar negocios adicionales bajo un mismo costo de operación.

5.1.2 CALIDAD Y CONFIABILIDAD DE DATOS

Las tareas manuales son sensibles a errores pudiendo asociar tasas de precisión cercana al 95% en actividades de clasificación y extracción manual de datos a través de teclado, desde el documento físico o desde una imagen, los que pueden ser totalmente evitados al implementar EDC con un amplio conjunto de validaciones contra bases de datos o listas predefinidas para garantizar la calidad de los datos, eliminando el riesgo asociado a la manipulación de datos sensibles, garantizando seguridad y confiabilidad sobre los datos capturados ya que no es posible que los operadores modifiquen la información contenida en las imágenes desde el sistema.

5.1.3 AHORROS EXTERNOS

Las evaluaciones pueden ser enviadas a la central de procesamiento desde diferentes localizaciones usando estaciones remotas de captura, evitando el envío de documentos físicos por servicios logísticos propios o de terceros a la vez que se reducen dichos costos y riesgos asociados.

5.1.4 CONCIENCIA ECOLÓGICA

Al incorporar tecnologías de la información para captura electrónica de datos como apoyo a procesos internos de negocios en cualquier tipo de organización, de manera intrínseca se está evitando la impresión hacia papel al usar impresoras virtuales causando así la reducción en la demanda de consumibles para impresoras físicas, papel y electricidad.

5.1.5 BENEFICIOS TANGIBLES E INTANGIBLES

Se demostró que la adopción de ABBYY FlexiCapture como tecnología de información para captura electrónica de datos aplicada al procesamiento de evaluaciones de talento humano genera ahorros económicos en un plazo menor a un año, y a pesar de requerir menos operadores lo cual es asociado a una práctica de reducción de personal, genera la oportunidad de mantener operadores de planta con un mejor salario y con mayor motivación al estar trabajando de manera moderna a diferencia de operadores contratados por horas según la demanda.

5.1.6 OPORTUNIDAD DE INTEGRACIÓN

Tal como se aprecian en las estadísticas del capítulo 1, tan solo el 16% extrae datos desde documentos para integrarlos directamente a procesos de negocio, lo que demuestra la existencia de un interesante mercado a la espera de soluciones globales para administración documental.

5.1.7 OPORTUNIDAD DE SERVICIOS BPO

Para organizaciones dedicadas a la consultoría de gestión de capital humano, quienes probablemente no cuentan con proyectos frecuentes de evaluación para sus clientes, pueden contratar a un aliado estratégico que ofrezca el procesamiento de evaluaciones como servicio, el cual les garantiza resultados precisos, confiables y entregados en períodos de tiempo muy cortos, convirtiéndolo en un atractivo para agrandar su cartera de clientes a la vez que se despreocupan de la inversión tecnológica en una solución EDC y su administración, pagando el costo de procesamiento por cada evaluación solamente cuando es requerido.

5.2 Recomendaciones

5.2.1 A LAS ORGANIZACIONES EN GENERAL

Dado que la resistencia al cambio es uno de los factores más comunes por los que se dilata o se evita la adopción de nuevas tecnologías perdiendo la oportunidad de mejorar procesos de negocio y reducir costos, la ejecución de una prueba de concepto en sitio es una alternativa muy viable que permite evaluar la calidad y oportunidad con la que implementan ABBYY FlexiCapture como solución para captura electrónica de datos inclusive integrada a repositorios o administradores documentales propios del cliente o propuestos.

5.2.2 CAMPO FUTURO DE INVESTIGACIÓN

Recomiendo para futuras investigaciones el analizar la factibilidad de implementar servicios BPO para captura electrónica de datos con infraestructura en la nube, que permitan a las organizaciones reducir su costo de propiedad en activos de tecnología de la información y contar con servicios para apoyo de procesos de negocio.

5.2.3 A LA MAESTRÍA

Integrar de manera más fuerte las cátedras durante el desarrollo de la maestría y profundizar en contenidos administrativos, financieros y gerenciales con mucha mayor orientación a tecnologías de la información.

5.2.4 A LA UNIVERSIDAD

Las universidades en general son el tipo de organizaciones que dependen en gran medida de la información radicada en documentos de papel. Recomiendo que aquellos proyectos de tesis prácticos y con oportunidad de ser aplicados en la Universidad, sean analizados con la finalidad de generar coyuntura para una relación comercial *ganar – ganar*, la misma que permita generar un caso de éxito para el crecimiento de los emprendimientos que realizamos los mismos estudiantes.

6. Bibliografía y Referencias

- [A] Análisis costo/beneficio de implementar tecnologías de captura electrónica de datos (EDC). Área de estudio: Evaluaciones al talento humano, Esteban Sansur Holguín, 2011.
- [B] ABBYY Partner Kit, ABBYY Software House, 2,011.
- [C] Mandato Constituyente número 8, dado y suscrito en Montecristi, Provincia de Manabí de la República del Ecuador, el 30 de abril de 2,008 y publicado en el Registro Oficial 592 el 18 de Mayo de 2,009.
- [D] Norma del Subsistema de Reclutamiento y Selección de Personal según acuerdo número MRL-2011-00142 del Ministerio de Relaciones Laborales del Ecuador.
- [1] <http://www.aiim.org/Research/Industry-Watch/Capture-and-Business-Process-2010>
- [2] <http://www.aiim.org/Research/Product-Studies/Capture>
- [3] <http://en.wikipedia.org/wiki/OCR-A>
- [4] Entrevista a Felipe Costales, Director de Proyectos de Investigación para Medición de Capital Humano en Sociométrika.
- [5] http://es.wikipedia.org/wiki/GNU_General_Public_License
- [6] <http://www.gnu.org/licenses/quick-guide-gplv3.html>
http://es.wikipedia.org/wiki/Software_libre_y_de_código_abierto
- [7] <http://quexf.sourceforge.net/>
<http://quexf.sourceforge.net/node/43>
<http://www.acspri.org.au/software>
- [8] <http://www.limesurvey.org>
- [9] <http://sqs2.net/>
<http://dev.sqs2.net/projects/show/sqs>
- [10] <http://www.cse.iitd.ernet.in/~aseth/udai/OMRProj/README.html>
<http://thegraffiti.wordpress.com/tag/omr/>
- [11] <http://latam.abbyy.com/empresa/>
<http://latam.abbyy.com/flexicapture/>
<http://www.abbyy.com/CaseStudies/>
<http://www.abbyy.com/Default.aspx?DN=2ea27609-a48e-4552-a35b-0df68f8e2516>

<http://www.abbyy.com/Default.aspx?DN=181bfe05-426c-4057-9c01-90be011e7e7e>

- [12] <http://www.aiim.org/community/product-guide/Capture/ABBYY-FlexiCapture>
- [13] <http://www.kofax.com/document-capture-software/>
<http://www.kofax.com/capture/features.asp>
<http://www.kofax.com/news/article.asp?id=1193>
<http://www.kofax.com/support/ascent/XtrataPro/4.5/faqs.asp>
<http://www.kofax.com/solutions/case-studies.asp>
<http://www.kofax.com/case-studies/the-iraqi-ministry-of-education-selects-kofax-intelligent-capture-and-exchange-to-automate-the-countrywide-processing-of-student-examinations/>
- [14] [http://www.aiim.org/community/product-guide/Capture/Kofax-Capture-\(Formerly-Ascent-Capture\)](http://www.aiim.org/community/product-guide/Capture/Kofax-Capture-(Formerly-Ascent-Capture))
<http://www.aiim.org/community/product-guide/Capture/Kofax-Transformation-Modules>
- [15] <http://www.irislink.com>
<http://www.irislink.com/c2-373-189/IRISCapture-Pro-8-5-for-forms.aspx>[http://www.aiim.org/community/product-guide/Capture/Kofax-Capture-\(Formerly-Ascent-Capture\)](http://www.aiim.org/community/product-guide/Capture/Kofax-Capture-(Formerly-Ascent-Capture))
- [16] <http://www.aiim.org/community/product-guide/Capture/IRIS-CapturePro>[http://www.aiim.org/community/product-guide/Capture/Kofax-Capture-\(Formerly-Ascent-Capture\)](http://www.aiim.org/community/product-guide/Capture/Kofax-Capture-(Formerly-Ascent-Capture))
- [17] <http://www.ubuntu.com/business/server/services>
https://www.redhat.com/rhel/purchasing_guide.html

7. Glosario

A

ADF	<p>Característica encontrada en escáneres, MFP's, fotocopiadoras y máquinas de fax en la que el alimentador de páginas contiene varias páginas y el dispositivo se encarga de tomar una página a la vez de manera automática.</p> <p>ADF es el acrónimo de Automatic Document Feeder.</p>
AJAX	<p>Técnica para desarrollar aplicaciones web interactivas, las cuales se ejecutan en el navegador del cliente, manteniendo una comunicación asíncrona con el servidor en segundo plano, lo que permite que las páginas puedan cambiar de contenido sin necesidad de ser cargadas iterativamente desde el servidor.</p> <p>AJAX es el acrónimo de Asynchronous JavaScript And XML.</p>
API	<p>Es un conjunto de procedimientos, funciones y métodos dentro de la programación orientada a objetos con la cual diferentes aplicaciones pueden comunicarse entre sí evitando que el usuario alimente al dispositivo de manera manual con una sola página a la vez.</p> <p>Es el acrónimo de Application Programming Interface.</p>
ASCII	<p>Código de caracteres basado en el alfabeto latino usado en el inglés moderno y otras lenguas occidentales el cual fue desarrollado en el ámbito de la telegrafía.</p> <p>Es el acrónimo de American Standard Code for Information Interchange.</p>

B

BPM	<p>Metodología que se enfoca en generar eficiencia mediante la gestión de procesos de negocio, los cuales se modelan, organizan, documentan y se optimizan de manera continua, permitiendo que la organización base su funcionamiento y los cambios en procesos en lugar de enfocarse en cambios funcionales</p> <p>BPM es el acrónimo de Business Process Management.</p>
BPO	<p>Define la subcontratación o tercerización con proveedores internos o externos de una o más actividades o funciones de procesos de negocio con la finalidad de generar una reducción del costo asociado al proceso subcontratado, con la finalidad de que el contratista concentre esfuerzos y estrategias a las actividades propias de su negocio que agregan valor.</p> <p>Entre las actividades que generalmente se delegan a BPO son la atención al cliente, la administración de talento humano, contabilidad y la administración tecnológica.</p> <p>BPO es el acrónimo de Business Process Outsourcing.</p>

D

DLL	<p>Archivos que contienen código ejecutable de computador los cuales son cargados a petición por un programa del sistema operativo. DLL es una extensión propia de este tipo de archivos para los sistemas</p>
-----	--

		operativos Windows. DLL es el acrónimo de Dynamic Link library.
	DMS	Sistema diseñado para rastrear, almacenar y controlar versiones de documentos electrónicos e imágenes digitalizadas de documentos físicos. Un DMS se considera ser una parte de un sistema ECM. DMS es el acrónimo de Document Management System.
	DPI	Unidad de medida que representa la cantidad de puntos contenidos en una pulgada lineal. Mientras más alta es esta medida, mejor será la calidad de imagen digitalizada así como también mayor el tamaño en bytes del archivo que lo contiene. DPI es el acrónimo de Dots Per Inch.
	Dúplex	Característica que permite a un escáner o un dispositivo multifuncional digitalizar una hoja de papel por ambas caras al mismo tiempo con el uso de dos lentes o sensores ópticos. Una forma de escaneo dúplex generalmente encontrada en dispositivos MFP es el mecanismo de reverso automático de papel, el cual se encarga de de dar la vuelta la hoja de papel posteriormente a haber digitalizado el anverso para proceder con la digitalización del reverso.
E		
	ECM	Sistema con el cual una organización se apoya para organizar y almacenar documentos y contenido en general relacionado con procesos de negocio. Un sistema ECM administra el ciclo de vida de los documentos y colabora para que los usuarios dependan de los documentos físicos para llevar a cabo procesos de negocio. ECM es el acrónimo de Enterprise Content Management.
	ERP	Sistema que integra la administración de la información interna y externa en toda una organización, involucrando los departamentos financiero, contable, manufactura, ventas y servicio y la relación con clientes. ERP es el acrónimo de Enterprise Resource Planning.
F		
	FTP	Protocolo de transmisión de archivos entre dos computadores usando internet. FTP es el acrónimo de File Transfer Protocol.
I		
	ISIS	Desarrollado por Pixel Translations (ahora EMC Captiva) en 1,990 como una interface estándar de la industria para la tecnología de digitalización de imágenes. ISIS es un estándar abierto para controlar dispositivos de digitalización el cual es aceptado por varias marcas fabricantes y aplicaciones de digitalización permitiendo al usuario controlar cualquier equipo de de digitalización usando una misma interface. ISIS se caracteriza por su velocidad de procesamiento de imágenes y poder ejecutar funcionalidades más complejas sobre una imagen escaneada.

		<u>ISIS es el acrónimo de Image and Scanner Interface Specificacion.</u>
G	GPL	Tipo de licenciamiento ampliamente utilizado por las aplicaciones de software libre, escrito originalmente por Richard Stallman. <u>GPL es el acrónimo de General Public License.</u>
K	KFI	Método de entrada en el que un operador digita manualmente los datos leídos que visualiza en la imagen en pantalla en lugar de leerlos en el documento físico. <u>KFI es el acrónimo de Key From Image.</u>
M	Metadata	<u>Estructura que describe datos acerca de los datos. Por ejemplo, sea el contenido de una fotografía los datos originales, la metadata de esta fotografía describiría la fecha y hora de la toma, el lugar, el autor, marca y modelo de la cámara de fotos utilizada entre muchos más.</u>
O	ODBC	Interface estándar que permite acceder a sistemas para administración de bases de datos, la cual es independiente de la base de datos en sí como del sistema operativo en donde reside. <u>ODBC es el acrónimo de Open Database Connectivity.</u>
P	PDF	Formato creado por Adobe Systems en 1,993 que permite mostrar un documento de manera independiente a la aplicación con que fue creado, el sistema operativo, hardware ó software. <u>PDF es el acrónimo de Portable Document Format.</u>
	POP3	Es un protocolo que usan clientes de correo electrónico el cual les permite recibir mensajes de correo que están almacenados en un servidor remoto hacia su bandeja de entrada en el cliente. <u>POP3 es el acrónimo de Post Office Protocol 3.</u>
Q	QWERTY	<u>Término con el que se asocia la distribución de letras en un teclado, el cual proviene de las primeras seis letras de la fila superior de teclas de dicho tipo de teclado.</u>
R	Rubber Band OCR	<u>Tecnología que permite al usuario dibujar un área rectangular que enmarca el texto deseado sobre el cual desea que el sistema realice el reconocimiento óptico de caracteres OCR ó el reconocimiento inteligente de caracteres ICR.</u>
S	Simplex	<u>Característica mínima de un escáner o un dispositivo multifuncional el cual digitaliza una sola cara de la hoja de papel.</u>
	SOAP	<u>Protocolo que permite la comunicación entre dos objetos de distintos procesos mediante el intercambio de datos en formato XML. SOAP es el acrónimo de Simple Object Access Protocol.</u>
T	Texto Plano	<u>Un formato de texto plano es aquel que contiene solamente</u>

		<p>caracteres de texto sin formato alguno, es decir, carecen de información respecto al tipo de letra, tamaño, etc., y puede ser abierto en cualquier editor de texto.</p> <hr/>
	TIFF	<p>Tipo de archivo de imagen el cual usa el algoritmo Huffman modificado para la compresión de datos. TIFF es el acrónimo de Tagged Image File Format.</p> <hr/>
	TWAIN	<p>Estándar lanzado en el año 1,992 cuya interface permite la adquisición de imágenes desde sistemas Microsoft Windows y Macintosh para que dispositivos de digitalización como escáneres y cámaras digitales puedan comunicarse con aplicaciones para procesamiento de imágenes. TWAIN no obedece a acrónimo alguno.</p> <hr/>
X		
	XML	<p>Conjunto de reglas que permiten la codificación de un documento para que pueda ser interpretado por un computador de tal modo que pueda ser de interpretación simple y usable en internet. XML es el acrónimo de eXtensible Markup Language.</p> <hr/>
W		
	WIA	<p>Es un modelo de controlador e interface de aplicación desarrollada por Microsoft para su sistema operativo Windows la cual estuvo disponible inicialmente en Windows ME en el año 2,000 y en sus posteriores sistemas operativos a la fecha. Permite que aplicaciones gráficas se comuniquen con dispositivos de imágenes como escáneres, cámaras digitales y equipo de video digital. Goza de una interface estandarizada generalmente independiente del modelo de dispositivo de imagen para manejar las funciones básicas de un dispositivo. WIA es el acrónimo de Windows Image Acquisition</p> <hr/>

8. Anexos

Anexo 1

8.1.1 MAPA DE CARACTERES ALFANUMÉRICOS DE TIPOGRAFÍAS RECOMENDADAS PARA ELEMENTOS DE TEXTO.

COURIER NEW

ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
0123456789

OCR - A

ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
0123456789

OCR - B

ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
0123456789

ARIAL

ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
0123456789

TIMES NEW ROMAN

ABCDEFGHIJKLMNOPQRSTUVWXYZ

abcdefghijklmnopqrstuvwxyz

0123456789

Anexo 2

8.1.2 CÓDIGOS DE COLORES PANTONE POSIBLES DE SER USADOS EN FORMAS PARA SEGREGACIÓN DE COLOR.

Tabla Pantone para segregación de color usando cualquier tipo de escáner.

Process Yellow U	100U	101U
102U	106U	107U
108U	109U	113U
114U	115U	116U
1205U	120U	1215U
121U	1225U	122U
123U	127U	1345U
134U	1355U	135U
1485U	148U	149U
1555U	162U	169U
372U	386U	387U
388U	3935U	393U
3945U	394U	3955U
395U	3965U	396U
461U	475U	503U
586U	587U	600U
601U	602U	603U
607U	608U	609U
614U	691U	705U
706U	712U	713U
714U	719U	Cool Gray 1U

Anexo 3

8.1.3 FORMA PARA SEGREGACIÓN DE COLOR.

FORMA DE EJEMPLO: SEGREGACIÓN DE COLOR ROJO

INSTRUCCIONES

1. Llene solamente con letra mayúscula tipo imprenta y números legibles, como en el ejemplo:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z 0 1 2 3 4 5 6 7 8 9

2. Marque su respuesta de selección así: No marque así:

En caso de haber marcado una opción incorrecta, corrija de la siguiente manera:

02 A B C D E *La selección inicial fue D, la que se ha tachado y se cambió por B.*

DATOS

Cédula:

Nombres:

Apellidos:

RESPUESTAS

01	A	B	C	D	E	16	A	B	C	D	E	31	A	B	C	D	E
02	A	B	C	D	E	17	A	B	C	D	E	32	A	B	C	D	E
03	A	B	C	D	E	18	A	B	C	D	E	33	A	B	C	D	E
04	A	B	C	D	E	19	A	B	C	D	E	34	A	B	C	D	E
05	A	B	C	D	E	20	A	B	C	D	E	35	A	B	C	D	E
06	A	B	C	D	E	21	A	B	C	D	E	36	A	B	C	D	E
07	A	B	C	D	E	22	A	B	C	D	E	37	A	B	C	D	E
08	A	B	C	D	E	23	A	B	C	D	E	38	A	B	C	D	E
09	A	B	C	D	E	24	A	B	C	D	E	39	A	B	C	D	E
10	A	B	C	D	E	25	A	B	C	D	E	40	A	B	C	D	E
11	A	B	C	D	E	26	A	B	C	D	E	41	A	B	C	D	E
12	A	B	C	D	E	27	A	B	C	D	E	42	A	B	C	D	E
13	A	B	C	D	E	28	A	B	C	D	E	43	A	B	C	D	E
14	A	B	C	D	E	29	A	B	C	D	E	44	A	B	C	D	E
15	A	B	C	D	E	30	A	B	C	D	E	45	A	B	C	D	E



FORMA DE EJEMPLO: SEGREGACIÓN DE COLOR GRIS

INSTRUCCIONES

1. Llene solamente con letra mayúscula tipo imprenta y números legibles, como en el ejemplo:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z 0 1 2 3 4 5 6 7 8 9

2. Marque su respuesta de selección así: No marque así:

En caso de haber marcado una opción incorrecta, corrija de la siguiente manera:

02 A B C D E *La selección inicial fue D, la que se ha tachado y se cambió por B.*

DATOS

Cédula:

Nombres:

Apellidos:

RESPUESTAS

01	A	B	C	D	E	16	A	B	C	D	E	31	A	B	C	D	E
02	A	B	C	D	E	17	A	B	C	D	E	32	A	B	C	D	E
03	A	B	C	D	E	18	A	B	C	D	E	33	A	B	C	D	E
04	A	B	C	D	E	19	A	B	C	D	E	34	A	B	C	D	E
05	A	B	C	D	E	20	A	B	C	D	E	35	A	B	C	D	E
06	A	B	C	D	E	21	A	B	C	D	E	36	A	B	C	D	E
07	A	B	C	D	E	22	A	B	C	D	E	37	A	B	C	D	E
08	A	B	C	D	E	23	A	B	C	D	E	38	A	B	C	D	E
09	A	B	C	D	E	24	A	B	C	D	E	39	A	B	C	D	E
10	A	B	C	D	E	25	A	B	C	D	E	40	A	B	C	D	E
11	A	B	C	D	E	26	A	B	C	D	E	41	A	B	C	D	E
12	A	B	C	D	E	27	A	B	C	D	E	42	A	B	C	D	E
13	A	B	C	D	E	28	A	B	C	D	E	43	A	B	C	D	E
14	A	B	C	D	E	29	A	B	C	D	E	44	A	B	C	D	E
15	A	B	C	D	E	30	A	B	C	D	E	45	A	B	C	D	E



Anexo 4

8.1.4 EJEMPLO DE FORMA BLANCO Y NEGRO CON BORDES DE TRAMA.

FORMA DE EJEMPLO: BLANCO Y NEGRO CON BORDES DE TRAMA

INSTRUCCIONES

1. Llene solamente con letra mayúscula tipo imprenta y números legibles, como en el ejemplo:
 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z 0 1 2 3 4 5 6 7 8 9

2. Marque su respuesta de selección así: No marque así:

En caso de haber marcado una opción incorrecta, corrija de la siguiente manera:

La selección inicial fue D, la que se ha tachado y se cambió por B.

DATOS

Cédula:


Nombres:

Apellidos:

RESPUESTAS

	A	B	C	D	E		A	B	C	D	E		A	B	C	D	E
01	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	16	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	31	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
02	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	17	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	32	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
03	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	18	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	33	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
04	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	19	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	34	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
05	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	20	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	35	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
06	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	21	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	36	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
07	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	22	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	37	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
08	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	23	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	38	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
09	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	24	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	39	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	25	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	40	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	26	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	41	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	27	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	42	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	28	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	43	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	29	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	44	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	30	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	45	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

INTELI8 Ecuador - www.inteli8.com



9 186276 466106

9. Tabla de Contenido

1. CAPTURA ELECTRÓNICA DE DATOS (EDC)	1
1.1 Introducción	1
1.2 Captura de Datos y Documentos	3
1.3 Captura de Documentos	3
1.3.1 Captura de Documentos por Importación	5
1.3.2 Captura de Documentos por Conversión	5
1.3.3 Captura de Datos desde Formas Electrónicas	6
1.3.4 Captura de Datos desde Formas	7
1.4 Clasificación de Soluciones según la Estrategia de Captura	7
1.4.1 Captura Ad-Hoc	7
1.4.2 Captura por Lotes	8
1.4.3 Captura Transaccional	10
1.5 Indicadores y Mercados	11
1.5.1 Escanear para Archivar contra Capturar para Procesar	14
1.5.2 Tercerización	15
1.5.3 Un vistazo a la Captura Manual de Datos	16
1.5.4 El Factor Humano	18
1.6 Mercados Verticales y Horizontales de Aplicación	19
1.6.1 Gobierno.....	19
1.6.2 Logística y Distribución.....	20
1.6.3 Educación	20
1.6.4 Banca y Finanzas.....	20
1.6.5 Legal	21
1.6.6 Seguros	21
1.6.7 Salud.....	21
1.7 Reconocimiento de Imágenes	22
1.7.1 Escáner	23
1.7.2 Diseño del Formulario	23
1.7.3 Procesamiento de Imagen.....	23

1.7.4	Reconocimiento Contextual	23
1.7.5	Resolución de Escaneo	23
1.7.6	Profundidad de Bit	24
1.7.7	Reconocimiento Óptico de Caracteres (OCR).....	25
1.7.8	Reconocimiento Inteligente de Caracteres (ICR).....	26
1.7.9	Reconocimiento Óptico de Marcas (OMR).....	27
1.7.10	OMR por medio de Lector Óptico de Marcas	28
1.7.11	OMR por medio de Escáner de Imágenes	30
1.7.12	Reconocimiento de Código de Barras (BCR)	32
1.8	Formas para Procesamiento por Computador	33
1.8.1	Tipos de Formas	34
1.9	Elementos de una Forma	36
1.9.1	Texto.....	36
1.9.2	Campo para Entrada de Datos	37
1.9.3	Campo de Selección	38
1.9.4	Grupo de Campos de Selección.....	38
1.9.5	Marcas de Referencia.....	39
1.9.6	Líneas Separadoras.....	40
1.9.7	Identificador de Forma	40
1.9.8	Imagen.....	40
1.9.9	Tabla	40
1.10	Consideraciones y Diseño.....	40
1.10.1	Formas para Segregación de Color	42
1.10.2	Forma Blanco y Negro con Fondos de Trama	43
1.10.3	Forma Blanco y Negro con Bordos de Trama	43
1.10.4	Forma Blanco y Negro Lineales	43
1.10.5	Selección del Diseño de Forma.....	43
2.	ANÁLISIS SITUACIONAL DEL ÁREA DE TALENTO HUMANO.....	45
2.1	Descripción, Importancia y Justificación de las Evaluaciones de Personal	45
2.2	Normativa Legal para el Ámbito Público [D]	47
2.3	Procesamiento Manual de Evaluaciones de Talento Humano [4].....	51
2.3.1	Descripción de Actividades.....	52

2.4 Debilidades Implícitas en el Procesamiento Manual	54
2.4.1 Formas de Respuestas	54
2.4.2 Tiempo de Procesamiento	55
2.4.3 Dificultad para Escalar	56
2.4.4 Errores en Ingreso de Datos	57
2.4.5 Archivo de Evaluaciones.....	57
2.4.6 Respecto a la Normativa Legal para el Ámbito Público.....	57
2.5 El Costo del Procesamiento Manual a Mayor Escala	59

3. SOLUCIONES PROPIETARIAS Y LIBRES PARA CAPTURA ELECTRÓNICA DE DATOS..... 61

3.1 Soluciones Libres y de Código Abierto	62
3.1.1 queXF versión 1.13.0 [7].....	64
3.1.2 Shared Questionnaire System versión 2.0 [9]	66
3.1.3 Udai OMR Tool [10]	67
3.2 Soluciones Propietarias	68
3.2.1 ABBYY FlexiCapture versión 10 [11]	68
3.2.2 Kofax Capture versión 10 [13]	73
3.2.3 IRISCapture Pro for Forms versión 8.5 [15]	77
3.3 Selección de la Solución EDC	81
3.3.1 Soluciones Libres y de Código Abierto	81
3.3.2 Soluciones Propietarias	82

4. IMPLEMENTACIÓN DE LA SOLUCIÓN PARA CAPTURA ELECTRÓNICA DE DATOS..... 85

4.1 Descripción del Proyecto	85
4.2 Tipos de Instalación	86
4.2.1 Instalación Independiente.....	87
4.2.2 Instalación Distribuida.....	87
4.3 Arquitectura Distribuida	88
4.3.1 Servidor de Aplicación	88
4.3.2 Servidor de procesamiento	88

4.3.3	Servidor de Licenciamiento	88
4.3.4	Servidor de Base de Datos.....	88
4.3.5	Estación de Configuración de Proyectos	89
4.3.6	Estación de Monitoreo y Administración	89
4.3.7	Estación de Digitalización	89
4.3.8	Estación de Verificación de Datos	89
4.3.9	Estación de Verificación	89
4.4	Descripción Funcional de la Solución	89
4.5	Configuración y Flujo de Trabajo de ABBYY FlexiCapture 10	90
4.5.1	Configuración	91
4.5.2	Procesamiento.....	93
4.6	Retorno Sobre la Inversión de TI (IT ROI)	104
4.6.1	Factores Considerados para el Cálculo del ROI	105
4.6.2	Cálculo del ROI	107
5.	CONCLUSIONES Y RECOMENDACIONES	109
5.1	Conclusiones.....	109
5.1.1	Productividad y Escalabilidad	109
5.1.2	Calidad y Confiabilidad de Datos.....	109
5.1.3	Ahorros Externos.....	109
5.1.4	Conciencia Ecológica	109
5.1.5	Beneficios Tangibles e Intangibles.....	110
5.1.6	Oportunidad de Integración.....	110
5.1.7	Oportunidad de Servicios BPO	110
5.2	Recomendaciones.....	111
5.2.1	A las Organizaciones en General	111
5.2.2	Campo Futuro de Investigación.....	111
5.2.3	A la Maestría	111
5.2.4	A la Universidad.....	111
6.	BIBLIOGRAFÍA Y REFERENCIAS	112
7.	GLOSARIO	114

8. ANEXOS.....	118
8.1.1 Mapa de caracteres alfanuméricos de tipografías recomendadas para elementos de texto....	118
8.1.2 Códigos de colores Pantone posibles de ser usados en formas para segregación de color.....	120
8.1.3 forma para segregación de color.....	121
8.1.4 Ejemplo de forma blanco y negro con bordes de trama.	123
8.1.5 Ejemplo de forma blanco y negro lineal.....	124
9. TABLA DE CONTENIDO	125