



**UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO**

FACULTAD DE INGENIERÍA

**SISTEMA COMPUTARIZADO PARA LA EVALUACIÓN DE
FORMULARIOS Y ALIMENTACIÓN AUTOMÁTICA DE LOS
RESULTADOS A UNA BASE DE DATOS**

T E S I S

PARA OBTENER EL TÍTULO DE:

**INGENIERO EN COMPUTACIÓN
PRESENTA:**

OSCAR ANZURES CALDERON

DIRECTOR DE TESIS CARLOS RIVERA RIVERA.

CODIRECTOR FERNANDO LEPE CASILLAS

MÉXICO. D.F.

Capítulo I

Antecedentes

La estadística es un valioso auxiliar de las distintas ciencias y técnicas; al servirles en la investigación de las leyes de comportamiento de ciertas variables que intervienen en un fenómeno cuyo valor, gracias a la Estadística ha sido puesto de manifiesto.

Conforme aumenta la necesidad de evaluar ciertas características de una sociedad o conjunto de personas aumenta también la necesidad de buscar mejores métodos que hagan más fácil la tarea, sobre todo cuando es muy extensa la cantidad de información por evaluar se hace más difícil el proceso de seleccionar y archivar la información, teniendo en cuenta además el tiempo para entregar los resultados. Por ello, se recurre a sistemas mecánicos o bien se ingenian métodos que agilicen ésta tarea, para el cual se considera también el aspecto económico.

1.1 Descripción de formularios para exámenes de opción múltiple

Un método muy común para evaluar, es de opción múltiple, es decir, para una pregunta se tiene un cierto número de opciones de respuesta, donde sólo una de ellas es la correcta o bien hay una o varias de interés. Esta forma de evaluación es muy práctica, ya que resulta más fácil de evaluar a cientos o miles de personas con tan sólo comparar las respuestas correctas con las proporcionadas por cada una de ellas, el personal dispuesto a evaluar puede no contar con un conocimiento acerca del tema. Un par de inconvenientes es que para una sola persona puede resultar tediosa esta tarea, sobre todo si se trata de cientos de exámenes, además de considerar el tiempo para entregar resultados. La forma más recurrida para evaluar es aquella en la cual se evalúa pregunta a pregunta de forma manual, la cual requiere de un conjunto de personas con los conocimientos necesarios para calificar los exámenes. Este método se complica aún más cuando el número de personas por evaluar aumenta, haciendo de esto una tarea mucho más tediosa.

Para el método de opción múltiple podemos mecanizar el sistema con el diseño de una plantilla que tenga una perforación en el espacio de la respuesta correcta, así se sobrepone la plantilla en el examen por evaluar y se cuentan los aciertos, finalmente el resultado se anota en una base de datos. De la misma forma se procede en cada examen.

Con la tecnología asociada a las computadoras de hoy en día, se ha simplificado considerablemente esta tarea. Así, se puede hacer uso de los lectores ópticos de marcas u OMR (*“optical mark reader”*). Un lector de marcas es una máquina que cuenta con pequeñas celdas fotoeléctricas que emiten luz a través de un canal, las cuales detectan si existe o no una marca en el impreso. Junto a esta máquina se cuenta con un conjunto de programas (*“software”*) que la controlan y manipulan la información obtenida. La velocidad con que se leen los formularios es de 8 a 10000 impresos por hora siendo muy confiable su evaluación.

Este lector óptico es una gran herramienta para todo tipo de evaluación, como por ejemplo; exámenes, encuestas, estadísticas, etc., pero tiene una limitante, que no todos tienen acceso a él, principalmente por el precio, ya que un dispositivo que pueda leer hasta 10000 impresos por hora cuesta alrededor de \$ 5000 dls. Sólo la industria o instituciones educativas con recursos propios y suficientes pueden adquirirlo. Esto obliga a que se recurra a los antiguos métodos de evaluación.

Como se mencionó anteriormente, el lector óptico reconoce si hay o no marca en el formulario, es decir, si la persona encuestada colocó una marca en el espacio indicado (óvalo sombreado). Esto quiere decir que el formulario debe ser especial ya que no todos los formularios pueden ser leídos en cualquier dispositivo lector, por lo cual dicho formulario es otro elemento con que se debe contar, además del dispositivo y el programa que lo controla, para la correcta evaluación.

En general, los formularios típicos contienen óvalos, que son los espacios designados para las respuestas, dependiendo de la encuesta que se vaya a realizar es el número de óvalos, los cuales están distribuidos según las necesidades. En la gran mayoría de los formularios los óvalos están en arreglo matricial, esto es, el formulario fue diseñado para que contuviera óvalos fáciles de localizar y ser leídos por la lectora; la razón es que el dispositivo tiene los fotosensores en línea, entonces al leer el formulario deben coincidir los óvalos, que tienen una posición única en renglón y columna, con los sensores del dispositivo.

Cada formulario debe contener marcas horizontales y verticales, renglones y columnas respectivamente, que ayuden a localizar la posición de los óvalos para su evaluación, cabe mencionar que estas marcas son pequeños rectángulos de color negro, y normalmente están distribuidas cerca de los contornos de la hoja. Aunque algunos otros formularios, dependiendo de las necesidades, las distribuyen en los espacios requeridos por dicho formulario. Todas las indicaciones y marcas auxiliares como sección de preguntas, aclaraciones, o datos adicionales, están de color rojo o verde, esto es porque los fotosensores del dispositivo no detectan estos colores, sólo pueden percibir el color negro, así, las marcas de posición y las respuestas sombreadas del encuestado serán las únicas detectadas por la lectora.

Algunas lectoras disponibles en el mercado pueden leer formularios con diseño por ambos lados lo cual aprovecha al máximo el espacio. Las velocidades con que actualmente se cuenta son de 8 a 10000 formularios por hora; el costo de cada dispositivo lector es proporcional a la velocidad de lectura. Para la evaluación de los formularios, la lectora tiene dos charolas de almacenamiento, una para la entrada y la otra para la salida (Fig.1). Los formularios deben estar apiladas en la charola de entrada y automáticamente el dispositivo extrae una a una para evaluarla depositándola posteriormente en la charola de salida.

En México sólo existe una empresa que se encarga del mercado de las lectoras ópticas, por lo que no hay mucha opción para escoger la mejor en cuanto a precios se refiere.

1.2 Software para ordenar la información obtenida por la lectora

El programa que acompaña a la lectora óptica es un sistema para manipular los datos obtenidos por la lectora, dicho programa organizará los datos de tal forma que sea fácil de interpretar.

El programa consta de un menú, donde se puede configurar el dispositivo y la información recibida por él. Por medio de este menú podemos indicarle a la lectora óptica que sólo tome información necesaria de las marcas laterales (color negro), por ejemplo de la marca número 1 a la 15 en renglón y de la marca número 1 a la 10 en columna.

Con este mismo menú podemos organizar la información, como por ejemplo, los títulos de las columnas, la extensión del archivo en donde será almacenada la información, etc. Una computadora está conectada a la lectora por medio del puerto serial, así la información proporcionada por la lectora es muy confiable y se puede depender de ella.

Finalmente, el programa es sencillo de manipular y cualquier persona no experta puede hacer uso de él y obtener buenos resultados.

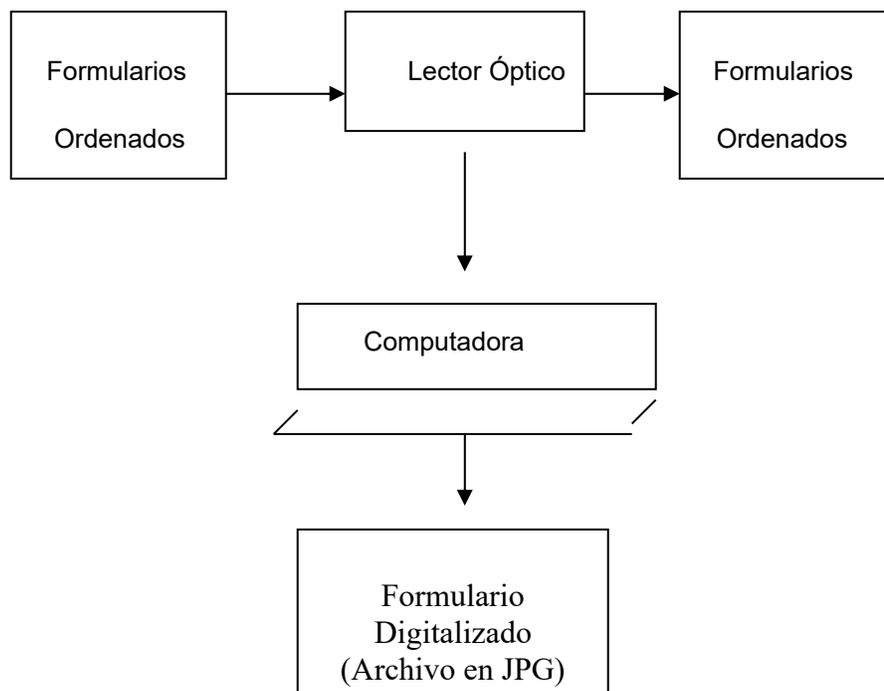


Fig. 1.1 Diagrama del lector óptico.

Capítulo II

Digitalización de imágenes

2.1 Digitalización de imágenes

Una imagen es una impresión visual que puede ser considerada como sinónimo de arte o bien como testimonio de lo acontecido. La digitalización de una imagen puede definirse como: la transformación óptica de la imagen en un simple arreglo de datos numéricos enteros, los cuales pueden ser manejados por una computadora, en donde cada valor del arreglo representa un *nivel* igual al nivel de luminancia existente o representado de la imagen original.

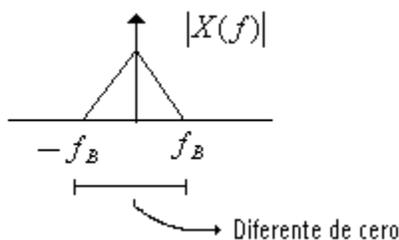
Para poder digitalizar una imagen se requieren dos procedimientos, la cuantización y el muestreo.

El muestreo de la imagen consiste en medir con algún dispositivo lector el valor de luminancia o nivel de gris en ciertas localidades de la imagen “(x,y)”. La cuantización organiza la información para que pueda ser representada en un arreglo numérico y dependiendo de la finalidad de la imagen si se aumentan estos dos valores se tendrá una mejor representación de la imagen original; considérese con ello que el tamaño del archivo para almacenarla es proporcional.

2.1.1 Teorema de Nyquist

Para poder llevar a cabo el muestreo de cualquier señal se debe cumplir el teorema de Nyquist:

Dada una función $x(t)$ con transformada de Fourier $X(f)$ que es de banda limitada

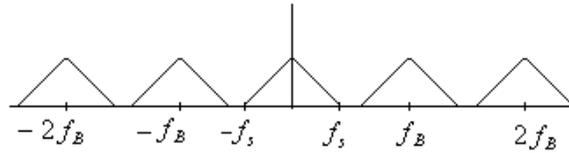


Si se toman muestras a una frecuencia

$$T = T_s, \quad f_s = \frac{1}{T_s} \quad \text{y} \quad f_s \geq 2f_B$$

$$x_n = x(nT_s), \quad n = \pm 1, \pm 2, \dots$$

La señal $x(t)$ se puede recuperar en forma exacta utilizando la secuencia x_n



Teorema de Nyquist en dos dimensiones

Dada una señal en un espacio de dos dimensiones $s(x,y)$ su transformada de Fourier es:

$$S(f_1, f_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(x, y) e^{-j2\pi(f_1x + f_2y)} dx dy$$

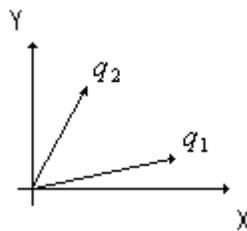
Esta señal es de banda limitada si:

$$S(f_1, f_2) = 0 \quad \text{para} \quad \begin{cases} |f_1| > f_{s1} \\ |f_2| > f_{s2} \end{cases} \text{ y}$$

Si se define una matriz de muestreo de la forma:

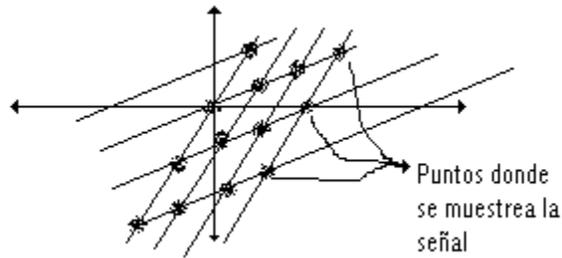
$$Q = [q_1 : q_2]$$

donde q_1 y q_2 son vectores no paralelos



entonces se puede formar una secuencia

$$S_{nm} = S(x_n, y_m) \quad , \quad \begin{bmatrix} x_n \\ y_m \end{bmatrix} = Q \begin{bmatrix} n \\ m \end{bmatrix}, \quad \begin{matrix} n = 0, \pm 1, \pm 2, \dots \\ m = 0, \pm 1, \pm 2, \dots \end{matrix}$$

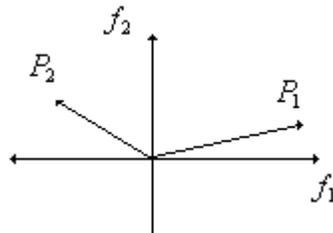


Asociada a la matriz Q hay una matriz P tal que

$$Q^T = P^{-1} \quad \text{donde}$$

$$P = [P_1 : P_2]$$

y P_1, P_2 son vectores en el dominio de la frecuencia.



El espectro de la señal muestreada está formado por una réplica infinita del espectro de la señal original centrada en todas las posibles combinaciones enteras de los vectores P_1 y P_2 .

La señal original se puede recuperar si no existe traslape de los espectros.

Una vez muestreada y cuantizada la imagen se representa matemáticamente por:

$$T = M \times N \quad (\text{renglones y columnas respectivamente})$$

$$G = 2^m; \quad G = \text{niveles de gris}$$

$$m = \text{número de bits para representar los niveles de gris}$$

La imagen es cuantizada en $M \times N$ puntos, cada uno de ellos es llamado pixel (abreviación de “*picture elements*”). Así, la cuantización representa el valor medido con un número entero por lo que M y N son valores enteros.

$$imagen = \begin{bmatrix} f(1,1) & f(1,2) & \dots & f(1,N) \\ f(2,1) & f(2,2) & \dots & f(2,N) \\ : & : & : & : \\ : & : & : & : \\ f(M,1) & : & \dots & f(M,N) \end{bmatrix}$$

Una vez que el formulario digitalizado está en un arreglo bidimensional, podemos continuar con el procesamiento digital de ésta, en el cual se puede alterar su tamaño, color, hacerla rotar en dos o tres dimensiones o bien hacer procesos con ella, como la restauración, realce, brillantez etc. Esto con el fin de adecuarla a nuestras necesidades.

El procesamiento digital de imágenes tiene un gran rango de aplicaciones, por ejemplo: la adquisición y restauración de imágenes recibidas por satélite, la transmisión de imágenes vía internet, en las imágenes médicas, radares, señales sonoras y acústicas, entre otras.

2.2 El explorador (escáner)

El explorador al igual que otros dispositivos tales como las cámaras digitales o bien las cámaras de video junto con un convertidor análogo digital, nos sirven para digitalizar una imagen. Estos dispositivos de alguna forma deben estar conectados a la computadora, la cual a su vez, necesita de un programa especial para recibir dicha información, procesarla y almacenarla.

El explorador o digitalizador de imágenes es un dispositivo que nos permite traer una imagen externa a una computadora. La información referente a las imágenes puede ser almacenada en discos duros, ópticos o bien flexibles para poder recuperarla cuantas veces sea necesario.

Cabe mencionar que el éxito para obtener una buena imagen depende no sólo del tipo de explorador y de los niveles de gris con que ha sido explorada, sino también de los programas con que se cuenta para el procesamiento de la misma, además de la impresora, que dará una buena calidad de imagen en papel. Si se tiene por ejemplo una impresora de matriz de puntos, la imagen impresa seguramente no será de buena calidad. Es por ello que para obtener una buena imagen impresa hay que tomar en cuenta la resolución de la imagen con que ha sido almacenada, así como el o los programas que se utilizaron para procesarla, el tipo de explorador y la impresora.

2.2.1 ¿Cómo funciona el explorador?

Un explorador consta de un bulbo fluorescente o incandescente, el cual ilumina en forma constante al ir explorando la imagen. La luz es reflejada por la imagen a un sensor fotográfico, que es normalmente un dispositivo acoplado por carga (CCD “*charged coupled devices*”) aunque a veces se usan fotodiodos. Ésta es una forma similar a como actúa el ojo humano, la luz de los objetos es reflejada y absorbida por los fotosensores del ojo, esta información es enviada al cerebro y finalmente asimilada. Los CCD’s son un conjunto de elementos fotosensibles que usualmente se encuentran ubicados en línea o bien en un arreglo bidimensional. Así, cada elemento produce un cierto nivel de voltaje que es proporcional a la cantidad de luz recibida. Posteriormente estos niveles de voltaje son procesados por un convertidor análogo digital (A/D) y finalmente almacenados como información digital en algún dispositivo.

La sensibilidad del explorador depende tanto de los sensores como del convertidor análogo digital, esto porque el convertidor puede cuantizar a diferentes niveles, por ejemplo un convertidor de un bit sólo rastreará un 1 o un 0, blanco o negro, si hay o no luz reflejada, si el convertidor es de 10 bits se tendrán 1024 tonos de gris (2^{10}), con 24 bits más de 16 millones de tonos (2^{24}) y así sucesivamente. Esta información es transmitida a la computadora por medio de una interfaz paralela, SCSI o bien por el puerto USB.

De igual forma se tiene el caso en las imágenes de color, sólo que aquí se manejan tres convertidores, uno para el color rojo, otro para el verde y otro más para el azul (RGB).

2.2.2 Sensores del explorador

Existen tres tipos de sensores: CCD, CMOS y PMT. CCD (“*charged coupled devices*”) y CMOS (“*complementary metal oxide semiconductor*”) son los sensores que usualmente tienen los exploradores de cama plana y las cámaras digitales, mientras que los PMT’s (“*PhotoMultiplier Tubes*”) son exclusivamente para los exploradores de tambor.

La resolución del explorador depende del número de elementos fotosensibles, entre más fotosensores tenga la resolución será mejor, aunque si se explora una imagen con una resolución menor, sólo se tomarán niveles de voltaje de algunos fotosensores y se descarta información dependiendo de la resolución dada. Por ejemplo, si se tiene un explorador de cama plana el cual contiene un arreglo de 400 CCD’s se puede decir que la resolución es de 400 spi (“*samples per inch*”), si se explora la imagen a 150 spi se descartará la información de algunos sensores.

Los sensores CMOS aparecieron en las cámaras digitales en 1996 y posteriormente en los exploradores de cama plana con alimentador a finales de 1997 en forma de CIS (“*contact image sensor*”). Una ventaja de los CMOS sobre los CCD’s es que los CMOS tienen los sensores y el convertidor análogo digital en un solo circuito mientras que los CCD’s lo traen por separado. Los PMT’s tienen la ventaja sobre los CCD’s y CMOS que estos últimos necesitan de cierto tiempo para calentarse y comenzar a explorar, por el contrario los PMT’s sólo amplifican la señal recibida y pueden explorar instantáneamente, aunque son mucho más lentos que los dos anteriores.

Es de importancia mencionar algunas diferencias que se han manejado en los libros de consulta respecto al ¿cómo se debe medir una imagen?. Cada autor utiliza las unidades de la resolución como mejor le parece. Aquí están algunas definiciones presentadas por el autor David Blatner;

“*Dot*”.- Esta unidad se maneja para la impresora, ya que ésta tiene una resolución en puntos por pulgada (“*dpi*”).

“*sample*”.- Se maneja para imágenes digitalizadas por medio de un explorador y aquí se habla de muestras por pulgada *spi*. Una muestra es una medida de un nivel de gris o bien de color en un lugar dado de la imagen. Aunque aquí está la confusión, porque algunos autores insisten en que la resolución de una imagen debe ser en *dpi*; pero la imagen en sí no tiene puntos sino muestras.

“*Píxel*” .- Para mostrar la resolución de una imagen se maneja *ppi*. Un *pixel* (“*picture element*”) es una coordenada (x,y) sobre un monitor, el cual puede de alguna forma representar los más de 16 millones de colores. Aunque algunos programas usan *ppi* en lugar de *spi* para la resolución de una imagen.

“*Res*”.- Algunos exploradores expresan la resolución en *res*. *Res* es una palabra para hacer referencia a las muestras por milímetro. “*Res 12*” quiere decir 12 muestras por milímetro el cual equivale a 304 *spi*.

Así, antes de que se digitalice una imagen con el explorador, se puede determinar con el programa proporcionado por el fabricante, las *spi* a las que se va a muestrear la imagen; entre mayor sean las *spi* a la que se muestrea, más lenta será la exploración; pero la calidad de la imagen será mejor.

Dependiendo de los fines con que se explora una imagen debe ser la resolución de la misma, es decir, si la imagen no necesita de una excelente calidad se puede explorar a 300 *spi* por ejemplo; pero si por el contrario la imagen necesita de una excelente calidad se debe dar una mayor resolución, teniendo en consideración que esto aumenta el tiempo de exploración y el tamaño del archivo.

Por último, podemos hacer mención de algunas tecnologías que sirven para la digitalización de imágenes o bien de información.

OMR - (“*Optical Mark Recognition*”, Reconocimiento Óptico de Marcas) OMR es la capacidad de detectar la presencia o ausencia de una marca en una hoja, y traducir esa información en un número o letra.

OCR - (“*Optical Character Recognition*”, Reconocimiento Óptico de Caracteres) Esta Tecnología permite reconocer caracteres impresos por una máquina y traducirlos a caracteres ASCII.

Código de barras – Esta tecnología consiste en leer e interpretar el código de barras por medio de un lector óptico.

Imágenes – Es la tecnología que reconoce caracteres escritos a mano y transforma la información a ASCII, para después administrar tanto las imágenes como los datos.

ICR - (“*Intelligent Character Recognition*”, reconocimiento inteligente de caracteres) Es la capacidad de reconocer caracteres escritos a mano (en letra de molde) y traducirlos a caracteres ASCII. Los formularios que contienen estos caracteres, pueden ser leídas por un lector, o recibidas a través de fax o directamente en un formato electrónico.¹

2.3 Formatos para almacenar imágenes

En la industria de la computación se han manejado estándares acerca de los formatos de las imágenes tales como GIF, TIFF, BMP, JPEG y PNG.

Cuando se habla del formato de un archivo se hace referencia al ¿cómo? ¿o de qué manera? ha sido guardada la información referente a ese archivo. El cómo ha sido guardada esa información también nos indica qué programa puede abrirlo y qué se puede hacer con ese archivo, como imprimirlo o bien editarlo. Por lo cual, el programa que manipula al explorador debe considerar el tipo de almacenamiento que se le puede dar a la imagen.

Los formatos BMP y TIFF son dos de los más utilizados en el diseño de presentaciones, mientras que los formatos JPEG, GIF y PNG son los más comunes en la Web.

El formato TIFF es el más usado en la industria, este formato puede ser codificado en nivel de gris, RGB o CMYK para imágenes a color. Puede tener niveles de gris de 8, 24, 32 (para CMYK) o 48 (para RGB) bits por muestra. Además puede ser almacenado con formato de compresión y sin compresión.

El formato GIF fue desarrollado en 1987 (“*GIF87a*”) y revisado en 1989 (“*GIF89a*”). Puede contener un máximo de 256 colores, los cuales son almacenados en una paleta que es parte propia del archivo GIF. Esto se llama índice de paleta de colores porque los colores en la imagen no son directamente representados por valores RGB.

El formato GIF usa una compresión sin pérdida, LZW², lo cual puede reducir el archivo sustancialmente.

¹Blatner David. Real World Scanning and Halftones, 2nd edition, Peachpit Press, 2000.

² La compresión LZW, permite comprimir identificando cuántas veces aparece un color o palabra en el archivo o imagen y lo representa con un número, así cada vez que aparezca esa palabra o color se sustituye por un número único, ahorrando mucho espacio al almacenar.

2.3.1 JPEG (Joint Photograph Experts Group)

El formato gráfico JPG, o más concretamente JPEG, fue desarrollado por el Joint Photograph Experts Group⁴, con él se trato de establecer un formato para intercambio en la red que soportara gráficos de gran calidad, muchos colores (24 bits de profundidad de color) y nivel de compresión elevado y configurable⁵.

Inicialmente este formato fue menospreciado por los diseñadores gráficos, pero poco a poco se abrió camino, hasta llegar a ser el formato más usado hoy en día al momento de transmitir gráficos de gran calidad.

Su sistema de compresión se basa en eliminar de la imagen aquellos colores que el ojo humano no puede distinguir y en agrupar gamas de colores similares. A mayor nivel de compresión, menor calidad en la imagen resultante. Casi todos los programas de creación de gráficos permiten ajustar al gusto del diseñador el nivel de compresión del archivo JPG a tratar, por lo que se debe experimentar con diferentes compresiones hasta llegar a un equilibrio entre peso en Kbytes de la imagen resultante y la calidad necesaria para una buena presentación.

Entre las ventajas de este formato gráfico destacan:

Poseer un sistema de compresión complejo y graduable, que permite una amplia gama de compresiones diferentes para una misma imagen. Además, el tamaño de la imagen comprimida no depende del número de colores de ésta, sólo de su nivel de compresión.

Maneja millones de colores (concretamente 16,777,216), por lo que resulta idóneo para fotografías e imágenes con contrastes y graduaciones de color.

Crear imágenes progresivas. Un JPG progresivo es la equivalencia en este formato de los GIF's entrelazados, el cual divide una imagen en una especie de persiana de franjas horizontales, de tal forma que al ser cargada por el navegador se muestran progresivamente series de franjas saltadas, lo que permite ver cómo se crea la imagen ante los ojos, a la vez que aumenta su definición en la pantalla. De esta forma el usuario no debe esperar a que la imagen se cargue en memoria completamente para poder visualizarla, sino que se observa poco a poco.

A pesar de estas ventajas, el formato JPG también tiene sus defectos, entre los que podemos citar:

Para grandes niveles de compresión, si la imagen es de gran resolución y muchos colores, para mantenerla con un aspecto óptimo se requieren ficheros gráficos de muchos Kbytes de memoria.

⁴ <http://www.w3.org/graphics/JPEG>

⁵ <http://adagio.music.northwestern.edu>

No permite la creación de imágenes de fondo transparente ni imágenes animadas. Para estos menesteres debemos recurrir al formato GIF.

Con pocos colores da archivos más extensos que el formato GIF, sin mejora apreciable de la calidad.

El formato PNG (*“Portable Network Graphics”*) entró en controversia sobre la compresión LZW usado para comprimir las imágenes GIF. Los archivos PNG tienen características semejantes a los formatos GIF: usa compresión sin pérdida, puede entrelazar imágenes, puede mostrarse en el navegador y puede usar paleta de colores de 256 colores. Sin embargo los archivos PNG tienen características muy importantes que los demás no las tienen y esto puede ser muy significativo sobre todo en la Web. Aunque no todos los navegadores pueden mostrar este tipo de archivos, sólo las últimas versiones de éstos tienen la capacidad de mostrarlos, por lo que se requiere nuevos *plug-ins* para actualizar los navegadores y puedan mostrar los archivos PNG.

Capítulo III

Características del sistema para evaluar los formularios

3.1 Medios utilizados para la evaluación

Para evaluar los formularios con el sistema diseñado, se cuenta con el siguiente equipo: una computadora con procesador Athlon a 1 GHz, 128 Mbytes de memoria RAM. Un *escáner* HEWLETT PACKARD 3400C con conexión al puerto USB (“*Universal Serial Bus*”). El “*software*” disponible es: sistema operativo Windows Millennium, MATLAB V.6 (entorno de programación), iGrafx Image V.1 (editor de imágenes), MsWord (editor de texto) y las hojas de respuestas que están diseñadas para operar con el sistema.

3.2 Diseño del formulario

El diseño del formulario es uno de los puntos más importantes para la evaluación, ya que debe tener algunas características importantes que lo definan perfectamente.

El formulario original con que se evaluó manualmente (antes) (Fig. 3.1), consta de una hoja en blanco tamaño carta (fotocopia), la cual tiene rectángulos como opción para cada pregunta, cada pregunta tiene seis rectángulos identificados con letras de la A a la F, el formulario consta de cuatro marcos distribuidos en la hoja, dos de los cuales son para contestar dos de los siete exámenes opcionales, los otros dos son para los exámenes obligatorios. En la parte superior derecha del formulario se encuentra un espacio para escribir el nombre del aspirante y en la parte inferior derecha del formulario se localiza un espacio para anotar la fecha de nacimiento y la rúbrica del aspirante.

A este formulario se agregaron marcas en la parte superior y en la parte lateral izquierda (Fig. 3.2), las marcas laterales indican la posición de las preguntas y las marcas superiores indican la posición de los <rectángulos de opción> de cada pregunta en los diferentes marcos. *Con este primer paso*, se logró hacer un reconocimiento del formulario para ubicar los <rectángulos de opción> de cada pregunta, todos los puntos donde se quería reconocer la marca estaban ubicados por medio de sus coordenadas en la imagen. Cada punto tenía una coordenada horizontal (‘y’) y una vertical (‘x’). Cuando se digitalizaron los formularios y se procedió a su evaluación algunos lo fueron correctamente; pero otros no, la razón es que si la hoja se desplazaba horizontal o verticalmente sobre el explorador, no se sabía dónde quedaban las marcas laterales y superiores del formulario; para corregir este detalle se debía establecer una marca que identificara al formulario en su posición correcta aún desplazándose sobre el explorador.

HOJA DE RESPUESTAS

NOMBRE _____

EXAMEN MATEMÁTICAS BÁSICAS						
1	A	B	C	D	E	F
2	A	B	C	D	E	F
3	A	B	C	D	E	F
4	A	B	C	D	E	F
5	A	B	C	D	E	F
6	A	B	C	D	E	F
7	A	B	C	D	E	F
8	A	B	C	D	E	F
9	A	B	C	D	E	F
10	A	B	C	D	E	F
11	A	B	C	D	E	F
12	A	B	C	D	E	F
13	A	B	C	D	E	F
14	A	B	C	D	E	F
15	A	B	C	D	E	F
16	A	B	C	D	E	F
17	A	B	C	D	E	F
18	A	B	C	D	E	F
19	A	B	C	D	E	F
20	A	B	C	D	E	F

EXAMEN INGENIERÍA ELÉCTRICA BÁSICA						
1	A	B	C	D	E	F
2	A	B	C	D	E	F
3	A	B	C	D	E	F
4	A	B	C	D	E	F
5	A	B	C	D	E	F
6	A	B	C	D	E	F
7	A	B	C	D	E	F
8	A	B	C	D	E	F
9	A	B	C	D	E	F
10	A	B	C	D	E	F
11	A	B	C	D	E	F
12	A	B	C	D	E	F
13	A	B	C	D	E	F
14	A	B	C	D	E	F
15	A	B	C	D	E	F
16	A	B	C	D	E	F
17	A	B	C	D	E	F
18	A	B	C	D	E	F
19	A	B	C	D	E	F
20	A	B	C	D	E	F

EXAMEN OPCIONAL DE						
1	A	B	C	D	E	F
2	A	B	C	D	E	F
3	A	B	C	D	E	F
4	A	B	C	D	E	F
5	A	B	C	D	E	F
6	A	B	C	D	E	F
7	A	B	C	D	E	F
8	A	B	C	D	E	F
9	A	B	C	D	E	F
10	A	B	C	D	E	F
11	A	B	C	D	E	F
12	A	B	C	D	E	F
13	A	B	C	D	E	F
14	A	B	C	D	E	F
15	A	B	C	D	E	F
16	A	B	C	D	E	F
17	A	B	C	D	E	F
18	A	B	C	D	E	F
19	A	B	C	D	E	F
20	A	B	C	D	E	F

EXAMEN OPCIONAL DE						
1	A	B	C	D	E	F
2	A	B	C	D	E	F
3	A	B	C	D	E	F
4	A	B	C	D	E	F
5	A	B	C	D	E	F
6	A	B	C	D	E	F
7	A	B	C	D	E	F
8	A	B	C	D	E	F
9	A	B	C	D	E	F
10	A	B	C	D	E	F
11	A	B	C	D	E	F
12	A	B	C	D	E	F
13	A	B	C	D	E	F
14	A	B	C	D	E	F
15	A	B	C	D	E	F
16	A	B	C	D	E	F
17	A	B	C	D	E	F
18	A	B	C	D	E	F
19	A	B	C	D	E	F
20	A	B	C	D	E	F

Fecha: 20 de Junio del 2001

Fecha de Nacimiento (dd/mm/año): _____

FIRMA DEL ASPIRANTE: _____

(En falta de la rúbrica inválida esta forma).

Fig 3.1 Hoja de respuestas de la DEPI

Otro problema que se suscitó con este diseño fue al girar el formulario en el explorador, ya que en la imagen obtenida las marcas de pregunta para los diferentes <rectángulos de opción> de una misma pregunta no coincidían, para un <rectángulo de opción> se tenía una coordenada en 'x' y otra en 'y', pero para otro <rectángulo de opción> de la misma pregunta que debía tener la misma coordenada en 'x' no la tenía. Para solucionar el problema, se debía identificar el ángulo de rotación y nuevamente girar la imagen para que las marcas de las preguntas coincidieran en los diferentes <rectángulos de opción>. Por último, si el formulario era una fotocopia, se considerarían los dos casos anteriores, primero que no se desplazara en forma horizontal y/o vertical al fotocopiar y segunda que no se rotara al obtener la fotocopia. En un caso extremo en el que se rotara y se desplazara el

formulario en la fotocopiadora y también que se rotara y desplazara al digitalizar el formulario en el explorador, sería muy difícil ubicar las marcas principales del formulario.

Además, como el nombre era el único identificador del formulario, cuando se evaluara cada uno de ellos se debería anotar el resultado uno por uno en una base de datos para saber de quién había sido el formulario evaluado.

HOJA DE RESPUESTAS

EXAMEN MATEMÁTICAS BÁSICAS

1	A	B	C	D	E	F
2	A	B	C	D	E	F
3	A	B	C	D	E	F
4	A	B	C	D	E	F
5	A	B	C	D	E	F
6	A	B	C	D	E	F
7	A	B	C	D	E	F
8	A	B	C	D	E	F
9	A	B	C	D	E	F
10	A	B	C	D	E	F
11	A	B	C	D	E	F
12	A	B	C	D	E	F
13	A	B	C	D	E	F
14	A	B	C	D	E	F
15	A	B	C	D	E	F
16	A	B	C	D	E	F
17	A	B	C	D	E	F
18	A	B	C	D	E	F
19	A	B	C	D	E	F
20	A	B	C	D	E	F

EXAMEN OPCIIONAL DE

1	A	B	C	D	E	F
2	A	B	C	D	E	F
3	A	B	C	D	E	F
4	A	B	C	D	E	F
5	A	B	C	D	E	F
6	A	B	C	D	E	F
7	A	B	C	D	E	F
8	A	B	C	D	E	F
9	A	B	C	D	E	F
10	A	B	C	D	E	F
11	A	B	C	D	E	F
12	A	B	C	D	E	F
13	A	B	C	D	E	F
14	A	B	C	D	E	F
15	A	B	C	D	E	F
16	A	B	C	D	E	F
17	A	B	C	D	E	F
18	A	B	C	D	E	F
19	A	B	C	D	E	F
20	A	B	C	D	E	F

Fecha: 20 de Junio del 2001
Fecha de Nacimiento (dd/mm/año):

NOMBRE

EXAMEN INGENIERIA ELÉCTRICA BÁSICA

1	A	B	C	D	E	F
2	A	B	C	D	E	F
3	A	B	C	D	E	F
4	A	B	C	D	E	F
5	A	B	C	D	E	F
6	A	B	C	D	E	F
7	A	B	C	D	E	F
8	A	B	C	D	E	F
9	A	B	C	D	E	F
10	A	B	C	D	E	F
11	A	B	C	D	E	F
12	A	B	C	D	E	F
13	A	B	C	D	E	F
14	A	B	C	D	E	F
15	A	B	C	D	E	F
16	A	B	C	D	E	F
17	A	B	C	D	E	F
18	A	B	C	D	E	F
19	A	B	C	D	E	F
20	A	B	C	D	E	F

EXAMEN OPCIIONAL DE

1	A	B	C	D	E	F
2	A	B	C	D	E	F
3	A	B	C	D	E	F
4	A	B	C	D	E	F
5	A	B	C	D	E	F
6	A	B	C	D	E	F
7	A	B	C	D	E	F
8	A	B	C	D	E	F
9	A	B	C	D	E	F
10	A	B	C	D	E	F
11	A	B	C	D	E	F
12	A	B	C	D	E	F
13	A	B	C	D	E	F
14	A	B	C	D	E	F
15	A	B	C	D	E	F
16	A	B	C	D	E	F
17	A	B	C	D	E	F
18	A	B	C	D	E	F
19	A	B	C	D	E	F
20	A	B	C	D	E	F

FIRMA DEL ASPIRANTE

(La falta de la cédula invalida esta forma)

Fig.3.2 Imagen de las marcas laterales y superiores del formulario

Esta forma de evaluar era también tediosa, porque se debía explorar la hoja de respuestas una por una y posteriormente con un programa hacer la evaluación; una vez que se tuviese la evaluación de cada examen (cuatro en el formulario) se debía escribir el resultado de cada uno. Aunque se reduce el tiempo de evaluación con respecto al método anterior

(evaluar pregunta por pregunta) se consideraba un trabajo tedioso, y más aún si aumentaba el número de formularios por evaluar.

Debido a lo anterior había que diseñar un nuevo formulario, que no presentara demasiados problemas para su análisis y a su vez no tuviera que desarrollarse un programa complejo que tomara mucho tiempo para evaluar el formulario.

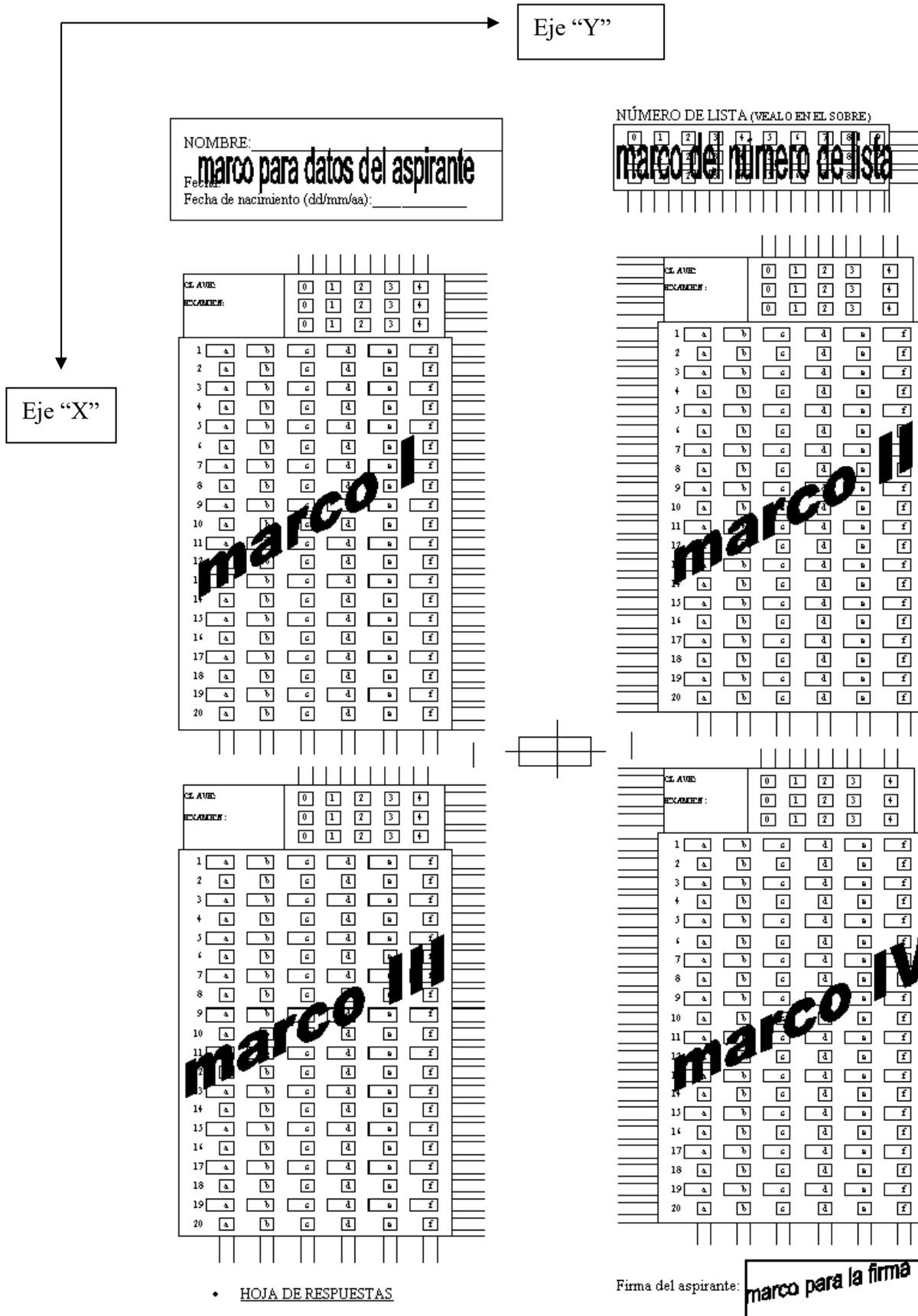
Con algunas correcciones a los puntos anteriores se diseñó un *nuevo formulario* que consta de un identificador para el alumno (número de lista), otro identificador para el examen por evaluar (clave del examen) y además se da opción a que el aspirante seleccione el orden de cualquiera de las cuatro áreas de respuesta. Los dos exámenes obligatorios (matemáticas y eléctrica básica) pueden ser contestados en cualquier marco, ya no necesariamente en los dos primeros del formulario. También se colocaron marcas para las preguntas y para los <rectángulos de opción> además de un <símbolo central> que ayuda a determinar si el formulario fue rotado o se desplazó horizontal y/o verticalmente al digitalizarlo.

Este nuevo diseño del formulario para responder los exámenes (Fig. 3.3a), es una impresión en blanco y negro en una hoja tamaño carta, en la parte superior izquierda tiene un <marco para los datos del aspirante> como son: nombre, fecha en la cual se presenta al examen y su fecha de nacimiento, en el orden día mes y año (dd/mm/aaaa). Frente a este marco (esquina superior derecha del formulario) existe el <marco de número de lista> del aspirante, este número se le da en el momento de presentarse al examen y consta de 3 dígitos. El <marco de número de lista> tiene pequeños <cuadros para rellenar>, son 10 cuadros alineados en dirección horizontal (renglón) numerados del 0 al 9 (columnas) y se tienen 3 de estos uno debajo del otro (3 renglones), así, tenemos un arreglo de 3 renglones por 10 columnas. Cada <cuadro para rellenar> está asociado a dos marcas en la parte lateral derecha (<marcas del nivel de significancia del dígito del número de lista>) y dos en la parte inferior del <marco del número de lista> (<marcas inferiores para identificar el dígito del número de lista>), las cuales sirven para identificar el número de columna ('y') en el formulario digitalizado. Por lo tanto, abajo del <marco del número de lista> se tienen 21 marcas (líneas verticales) y en la parte lateral derecha 6 (líneas horizontales)(Fig. 3.3b).

En el formulario se encuentran cuatro marcos para las respuestas de los cuatro exámenes por evaluar, cada uno de estos marcos tiene un arreglo de <rectángulos de opción> de dimensiones: 20 x 6 (Fig.3.3c), ya que son 20 preguntas con 6 <rectángulos de opción> para cada una de ellas. Los <rectángulos de opción> están identificados con las letras 'a', 'b', 'c', 'd', 'e' y 'f'. En la parte inferior de cada marco se desprenden las <marcas de los rectángulos de opción>, dos marcas por cada <rectángulo de opción>, una indica el inicio y la otra el final de éste; en la parte lateral derecha de los marcos I y III se desprenden también 2 marcas por renglón (<marcas de pregunta>) y para los marcos II y IV (inferior y superior derecho) las <marcas de pregunta> se ubican en su parte lateral izquierda.

En la parte superior de cada marco hay un <marco de dígitos de clave> para escribir la clave del examen en cifras decimales ("números del 0 al 9") esta clave consta de 3 dígitos, también hay un espacio para escribir el nombre del examen con letra de molde y frente a éste se localizan los cuadros para rellenar el número de la clave antes escrita; el primer "renglón" es para el dígito más significativo y el tercer renglón para el menos significativo.

El <símbolo central> del formulario (rectángulo con una cruz encima) es el que indica si la hoja está rotada o en posición adecuada para su evaluación. Las <marcas laterales al símbolo central> (líneas verticales) son para ubicar la coordenada 'x' de cada línea horizontal que se desprenden de cada marco. Hay diferentes alteraciones del formulario que pueden hacer que el sistema no evalúe correctamente. Debido a lo anterior sólo se deben rellenar los espacios indicados.



• HOJA DE RESPUESTAS

Firma del aspirante:

Fig. 3.3a Descripción del formulario

Marco para los datos del aspirante

NOMBRE: _____
Fecha: Fecha de nacimiento (dd/mm/aa): _____

Marco para el número de lista del aspirante

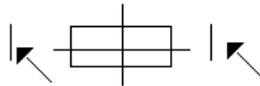
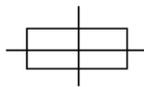
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

Below the table are 10 vertical tick marks, each aligned with a column of the table above.

Marcas del nivel de significancia para los dígitos del número de lista del aspirante.

Marcas inferiores para determinar el dígito del número de lista del aspirante

Símbolo central del formulario



Marcas laterales al símbolo central

Marco para la firma del aspirante

Firma del aspirante:

--

Fig. 3.3b Descripción del formulario

Finalmente, en la parte inferior derecha del formulario se indica que el aspirante firme para constatar que fue él quien contestó el examen y no exista un posible fraude. La firma estará contenida dentro del marco indicado (<marco para la firma del aspirante>) y no deberá salir de éste para evitar marcas indeseables; en el lado inferior izquierdo del formulario se localiza la leyenda: “HOJA DE RESPUESTAS”.

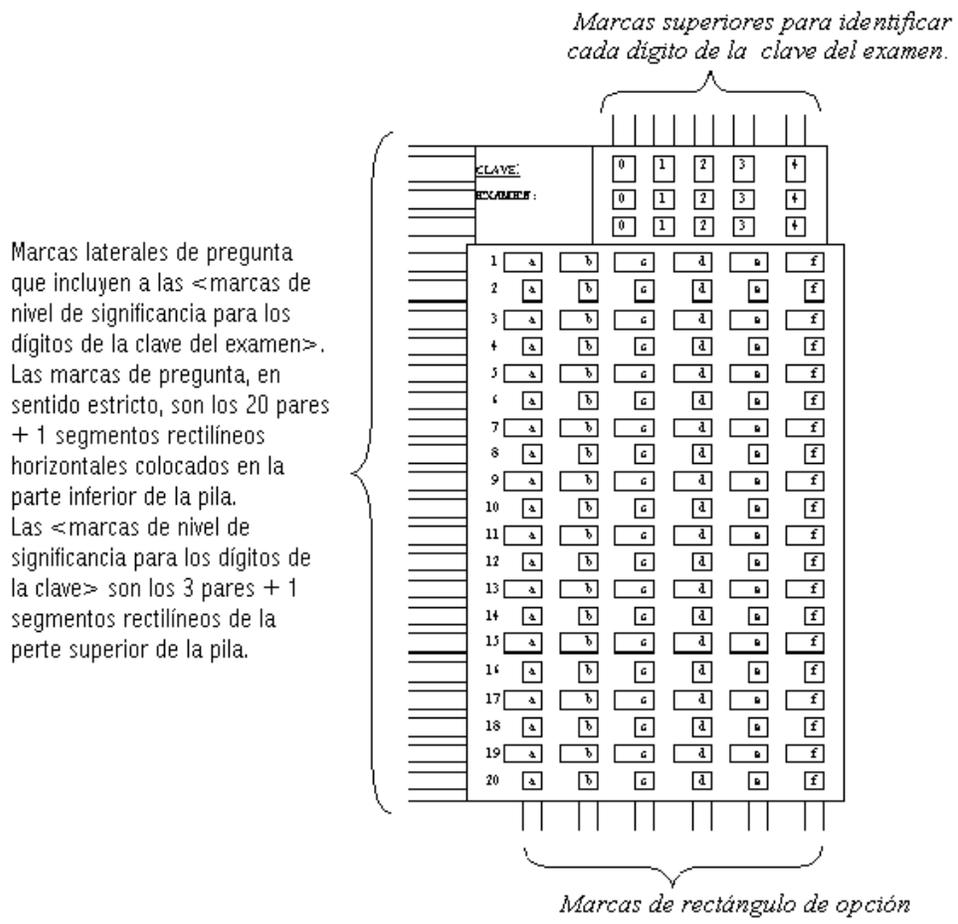


Fig. 3.3c Descripción gráfica del formulario.

3.3 Descripción del sistema

Una vez que el aspirante respondió el examen, se procede a su evaluación, para ello se requiere que el formulario se digitalice con el explorador y que la imagen obtenida sea a 75dpi como dibujo en blanco y negro. Al finalizar la digitalización del formulario, se debe almacenar en la computadora con formato JPG y factor de compresión 1. Este archivo deberá guardarse en el mismo directorio en donde se encuentra el programa fuente para la evaluación.

El nombre del archivo que tendrá cada imagen deberá estar numerado en la parte final del nombre, es decir, sólo variarán los últimos dígitos. Por ejemplo, la imagen uno se llamaría: “a1.JPG”, la imagen dos: “a2.JPG” y así sucesivamente. Cabe mencionar que el programa

MATLAB hace distinciones entre mayúsculas y minúsculas, lo que debe considerarse al momento de almacenar la imagen.

El enumerar las imágenes es con el fin de automatizar la evaluación, ya que el programa para evaluar los formularios sólo toma el nombre proporcionado por el operador del sistema y lo varía en el último dígito, aumentándolo de uno en uno. Este requisito (numerar las imágenes) es necesario sólo si se desea evaluar un conjunto de formularios, ya que en el caso de la evaluación individual las características del nombre no importan, la imagen sólo deberá estar con formato JPG.

Cuando ya se tiene a todos los formularios digitalizados y almacenados como imágenes con formato JPG, se ejecuta el programa MATLAB y se posiciona en el directorio donde se encuentra el programa fuente para la evaluación, posteriormente se ejecuta el programa *calificar* en la línea de comandos y aparece un menú, el cual se describe a continuación:

Primero aparece un cuadro de diálogo en el que se pregunta si desea evaluar 9 ó 10 áreas de conocimiento diferentes, la razón de esto es porque se da opción para que el sistema pueda evaluar hasta 10 exámenes de áreas diferentes, aunque actualmente en la Sección de Ingeniería Eléctrica de la División de Estudios de Posgrado sólo se evalúan 9.

Al responder la pregunta anterior el programa presenta un menú en el que se deben introducir las claves de los exámenes por evaluar así como el valor de las respuestas correctas e incorrectas; una vez introducidos estos datos se muestra otro menú con 7 botones de opción, los primeros 5 son para introducir las respuestas correctas de los diferentes exámenes, la siguiente opción es: *calificar exámenes*, que sirve para iniciar la evaluación y la última opción del menú es: *salir*, que finaliza el programa de evaluación.

La opción *proporcionar respuestas correctas* consta de una “ventana” que contiene opciones de tipo radio, las cuales son pequeños círculos que pueden ser seleccionados ya sea con el botón izquierdo del ratón o bien con la barra espaciadora del teclado⁶. Esta “ventana” es de aspecto similar al formulario, la diferencia es que no tiene espacio para el nombre, fecha y la firma del aspirante.

A cada “radio” de respuesta se le asoció una variable binaria, que posteriormente servirá para hacer la comparación de las respuestas de los aspirantes con las ingresadas en este formulario virtual. Recuérdese que no se deben borrar estas variables, pues si se llegaran a borrar no se tendrá ninguna evaluación porque no habrá un patrón de comparación. Si se desea almacenar las respuestas correctas en un archivo sólo escriba en el editor de comandos de MATLAB “save” y enseguida el nombre del archivo en el que se desee guardar las respuestas correctas.

Para cada pregunta se tienen 6 opciones de tipo radio que corresponden a los <rectángulos de opción> de cada marco del formulario. A cada opción se le asigna una variable, se tendrán entonces 6 variables por pregunta. Para hacer más sencilla la operación de

⁶ MATLAB presenta una opción para que se ingresen datos en forma gráfica.

comparar las respuestas correctas con las proporcionadas por el aspirante las 6 variables de las opciones de tipo radio se unen en un vector identificado con un solo nombre.

En MATLAB la comparación de matrices y vectores requiere de una sola instrucción para llevarse a cabo. Es por ello que toda la información obtenida de la digitalización de los formularios se maneja en vectores o matrices.

Al finalizar la evaluación de los formularios, se debe revisar, en el directorio donde está el código fuente del programa, si existe el archivo "errores.txt", en este archivo se almacenan los errores que se presentaron durante la evaluación. Si este archivo no aparece, quiere decir que no hubo ningún error.

Una vez corregidos los errores si se presentaron, se tiene la opción de guardar las calificaciones obtenidas (archivo "*calificaciones.txt*") en una base de datos. Por último, se da la opción de graficar las estadísticas de los exámenes presentados, lo cual permite observar la gráfica del número de aspirantes que respondieron correctamente a cada una de las preguntas de los diferentes exámenes, es decir, todos los exámenes que se presentaron tendrán su gráfica (No. de pregunta vs personas que respondieron correctamente), con éstas gráficas se evaluará el nivel general de conocimiento de los aspirantes, no se evaluará individualmente. Para que esta opción esté disponible (graficar), se deberá tener en la base de datos: las claves de los exámenes y el nombre de la materia en la tabla correspondiente, ya que el programa abrirá la base de datos y obtendrá el nombre de la materia para mostrarlo como título. Por su parte, las tablas: *estadísticas* y *presentaron*, deberán contener los datos del número de alumnos que presentaron el examen y el número de respuestas correctas de cada una de las preguntas, todos estos datos se almacenarán automáticamente al momento de evaluar.

Capítulo IV

Algoritmo para la evaluación del examen

Se usará la siguiente notación para las <marcas de pregunta> y para las <marcas de rectángulo de opción>:

Sea $[x_1, x_2]$ un intervalo cerrado de valores enteros y $x_2 \geq x_1$;

sea $[y_1, y_2]$ un intervalo cerrado de valores enteros y $y_2 \geq y_1$.

A cada marca vertical en el formulario le corresponden los índices $([x_1, x_2], y)$ y a cada marca horizontal le corresponden los índices $(x, [y_1, y_2])$.

Cuando el formulario por evaluar se encuentra almacenado como imagen con formato JPG, se abre el archivo de la imagen con el programa MATLAB y se almacena en una variable matricial la información como un arreglo de datos, dicho arreglo contiene elementos discretos entre 255 y 0, el número 255 representa el color blanco y el 0 el color negro. Otra característica importante del arreglo es la dimensión: 877 x 638 (*renglones y columnas*); éstos valores son debidos a la resolución dada a la imagen al momento de digitalizarla (“75 dpi”). Por lo anterior, si se altera la resolución al explorar el formulario, las dimensiones cambian y por tanto el sistema no evaluará correctamente.

Dentro de la matriz de valores discretos se definirán pequeñas matrices “RO” de dimensiones variables, estas matrices “RO” corresponderán a los <rectángulos de opción> de cada marco del formulario. Aunque todas las <marcas de las preguntas> y las <marcas de los rectángulos de opción> deberían estar situados a una distancia igual en pixeles (misma dimensión para todas las matrices “RO”) dentro de la matriz del formulario, no lo están, debido a la distorsión que presenta el explorador al momento de digitalizar el formulario. Las dimensiones de estas matrices las determinan, tanto las <marcas de preguntas> como las <marcas de los rectángulos de opción>. Como cada <rectángulo de opción> tiene asociadas 4 marcas, 2 en la parte inferior del marco que indican el principio y fin del <rectángulo de opción> y las otras 2 en la parte lateral derecha o izquierda según el marco que se analice, se tiene entonces delimitado al <rectángulo de opción> dentro de la matriz del formulario digitalizado. Así, se obtienen las dimensiones de cada matriz con el número de pixeles que ocupa el <rectángulo de opción> en renglón y columna dentro de la matriz del formulario.

El promedio de los valores discretos de la matriz “RO” determinará si el aspirante sombreó o no alguna opción de la pregunta en el formulario.

Aunque los <rectángulos de opción> de cada marco del formulario son determinados por sus coordenadas en la matriz del formulario digitalizado, existe una distorsión debida al explorador, que altera a la imagen y por lo tanto a las <marcas de pregunta> y a las <marcas de rectángulos de opción> dibujadas en él; por ejemplo, aparecen puntos negros

que no pertenecen al formulario o por el contrario, no están los que deberían estar, si el explorador falla puede alterar la imagen y estirarla en ciertas zonas. También, existe ruido por causa externa al explorador; por ejemplo, si al imprimir el formulario aparecieron puntos de más debido a la mala calidad de la impresora, defectos de la hoja blanca, etc. Todo este ruido y distorsión que puede haber en la imagen lo considera el sistema para evaluar los formularios; pues un caso extremo se da cuando la hoja se mancha accidentalmente, o bien la imagen se distorsiona: por rotar el formulario al explorar, o al obtener una fotocopia del original girada ya sea por la máquina fotocopidora o por el fotocopista, etc. Cualquiera de estas situaciones extremas será indicada por el programa en el archivo “errores.txt” con el número asociado al error y mostrará una breve descripción del tipo de error, además del nombre de la imagen en el que se presentó. Por el contrario, si el archivo no apareció al finalizar la evaluación de todos los formularios, quiere decir que no hubo error alguno y se considera confiable la evaluación.

En caso de que exista algún error al evaluar, el formulario con error no se evaluará hasta que se haya corregido, o en el mejor de los casos hasta que se explore el formulario nuevamente; el sistema continuará con la evaluación de los otros formularios e indicará en el archivo de errores generado, el error correspondiente. El sistema guiará al usuario sobre el manejo de las diferentes versiones de este archivo como se verá mas adelante.

En el centro del formulario existe un pequeño rectángulo con unas líneas en cruz que lo dividen en cuatro partes (<símbolo central>). Este diseño tiene que ver con la ubicación del centro para determinar si el formulario se giró o desplazó sobre la cama del explorador al digitalizarlo.

El <símbolo central> es el primer símbolo que analiza el programa (Fig. 4.1). Como sólo existe esta marca en el centro del formulario, se ubica el centro del formulario con sólo determinar el promedio de los valores en los distintos renglones de cada columna ($\forall y, V(y) = \sum_{x=1}^{877} f(x, y)$) de la imagen y así se obtiene un arreglo unidimensional de 1 x 638, el valor promedio para el <símbolo central> es 477 y se encuentra entre el elemento 300 y 370 del vector, este intervalo es debido al posible desplazamiento sobre el explorador. Si esta imagen ha sido desplazada, ya sea en la impresora, en el explorador o en la fotocopidora, el programa localiza el centro sin problemas. Esto sólo si se corrió en dirección horizontal o vertical, no incluye la rotación.

Si aparece un punto imprevisto exactamente en el centro de la imagen, se procede a cambiar de coordenada ‘y’ de búsqueda para determinar si se trata de un error por un punto o bien de un error por rotación. El error por un punto puede ser corregido por el programa y el error por rotación no puede ser corregido ya que al girar matemáticamente la imagen a su posición correcta las <marcas de preguntas> y las del símbolo central se deforman, y en este último caso el error aparecerá en el archivo “errores.txt” con su número descriptor del tipo de error y el formulario se descartará para evitar que se evalúe en forma incorrecta.

El <símbolo central> del formulario indica si la imagen está o no rotada (Fig. 4.2). Si la imagen está rotada aparecen dos líneas paralelas y casi colineales con sus puntos extremos más cercanos unidos con un segmento rectilíneo, pero con un pixel de diferencia en la

coordenada 'x', los extremos unidos definen un escalón. Aunque también se presenta el escalón en dirección vertical, el análisis del programa para evaluar no lo considera; pues lo analiza horizontalmente. Por el contrario, si la imagen no está rotada sólo aparece una línea recta y por lo tanto las coordenadas de fila 'x' son constantes en todas las líneas horizontales del <símbolo central> de la imagen. Nótese en la figura (4.2) que aún cuando el formulario está rotado, por efecto del explorador el escalón se ve en ángulos rectos. Si el ángulo de giro fuese mayor, aparecerían más líneas horizontales.

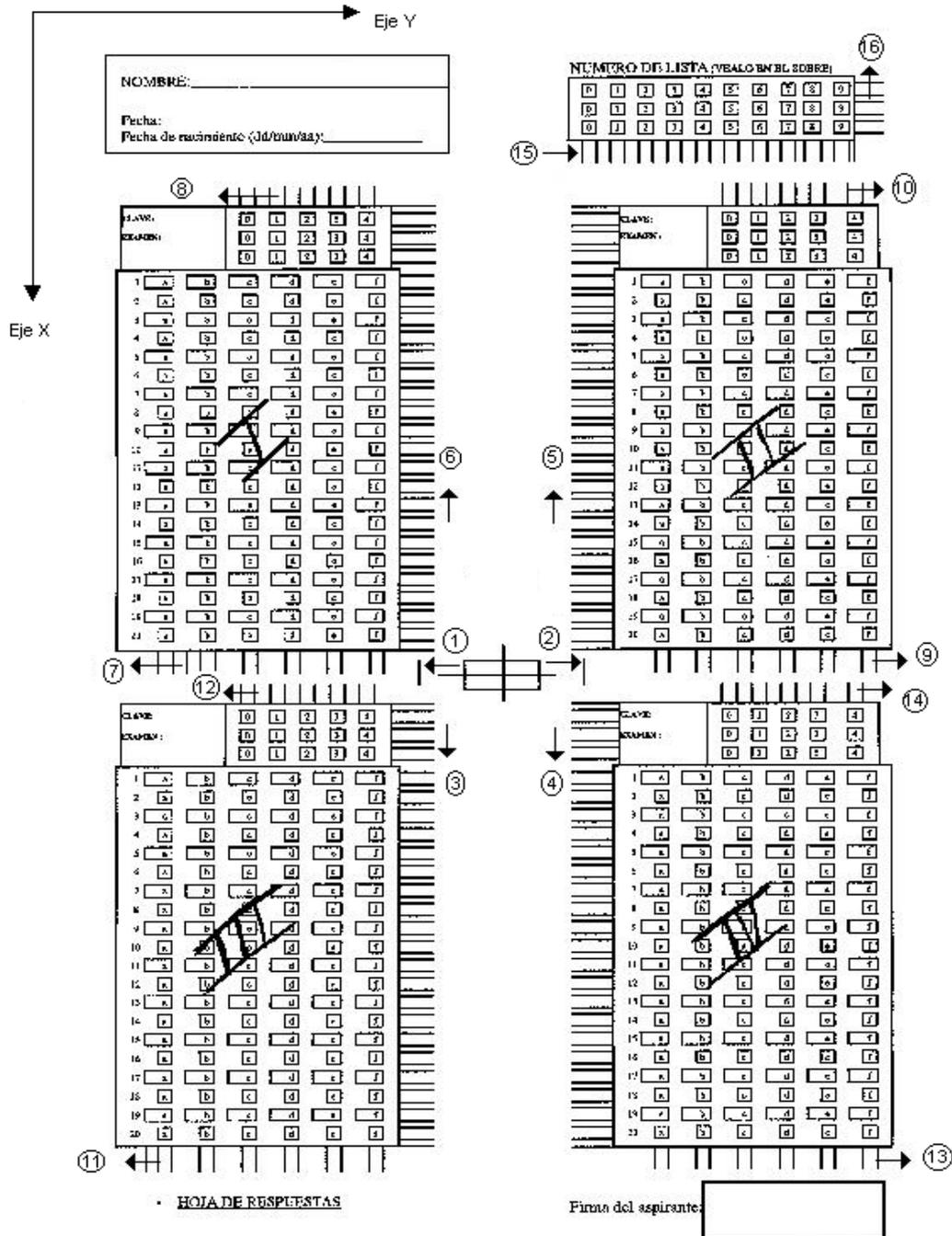


Fig 4.1 Orden en que se identifican los <<vectores de la matriz completa>> $(x, [y1, y2, y3])$ y/o $([x1, x2, x3], y)$ que corresponden a los distintos segmentos rectilíneos del formulario.

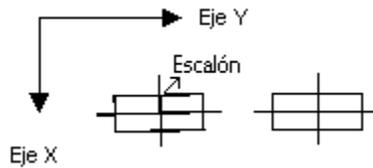


Fig. 4.2 En la parte izquierda de la figura se encuentra el <símbolo central > rotado, en la parte derecha se muestra el <símbolo central> en su posición correcta.

Una vez localizado el centro del formulario en la imagen, el programa analiza las <marcas laterales> al <símbolo central> (ver figura 3.3b), que sirven para localizar la posición de las <marcas de las preguntas> de los cuatro marcos por evaluar. Estas <marcas laterales> están ubicadas de tal manera que al localizar el <símbolo central> del formulario puedan ser ubicadas con su coordenada 'y' e intervalo de valores 'x'. La coordenada 'y' define una recta vertical que permite localizar las <marcas de las preguntas>. Estos últimos son segmentos rectilíneos horizontales, que cuando están en los marcos II y IV, se intersectan con la prolongación rectilínea de la <marca lateral derecha> del <símbolo central>. El mismo procedimiento se lleva a cabo para ubicar las <marcas de las preguntas> de los marcos I y III; pero se utiliza como referencia vertical la <marca lateral izquierda> del símbolo.

La recta definida por $y = \text{constante}$, que corresponde al número de columna de la <marca lateral izquierda>, es la primera en ser localizada y si no hay ninguna alteración en ella, se almacena su coordenada 'y' y se procede a localizar las <marcas de preguntas> del marco III. Obsérvese en la fig. (3.3a) que en la dirección vertical definida por la <marca lateral izquierda> y por debajo de las <marcas de pregunta> conectadas al marco III, existe un espacio completamente en blanco (espacio al lado derecho del letrero HOJA DE RESPUESTAS), como no hay puntos negros en dicho espacio y el marco III del formulario tiene conectadas 48 <marcas de preguntas>, si aparecen más de 48 marcas en la imagen se concluye que el formulario está rotado 180°. Si la imagen está rotada así, aparecerá el <marco de número de lista> en la parte inferior izquierda del formulario, debido a este error, que se indica en el archivo "errores.txt", el programa para evaluar descartará el formulario y se deberá explorar nuevamente. Un formulario rechazado no deja rastros en los archivos generados por el sistema, excepto el registro en el archivo de errores, donde se identifica al formulario rechazado por su nombre secuencial.

Los errores por puntos aislados aparecen en todo el formulario en posición aleatoria. El programa para evaluar lo considera durante la identificación de las <marcas de pregunta> de cada marco y para el <símbolo central> del formulario. En la prolongación de cada una de las dos <marcas laterales> del símbolo central deben encontrarse 48 puntos negros, ni más ni menos. Si no ocurre así, se atribuye a un error por punto, y para atacar el problema, se modifica ligeramente el valor de la coordenada 'y' hasta quedar satisfecho con 48 puntos exactamente.

Si la imagen del formulario no está rotada y el número de <marcas de pregunta> en el marco III es 48, se procede al reconocimiento del marco IV. Recuérdese que en este trabajo

con <marcas de pregunta> se designa propiamente a las 40 marcas de pregunta y también impropriamente, a 8 marcas más para el nivel de significancia de la clave del examen. Para el marco IV se toma el valor de la coordenada 'y' de la <marca lateral derecha> del símbolo, se prolonga esta recta y se obtiene cada 'x' que indica la posición de cada fila de las <marcas de pregunta> de este marco.

Al igual que el marco III, el marco IV tiene también 48 <marcas de pregunta> sólo que en la parte lateral izquierda al marco, y si existe alguna alteración en alguna de ellas se presentará un error.

Una vez determinado el número de fila de las <marcas de preguntas> en el marco IV, se analiza el marco II. En la parte superior del marco II se localiza el marco que encierra al número de lista (<marco de número de lista>), esto se considera al momento de almacenar la posición de las <marcas de las preguntas> del marco II ya que se deben diferenciar los pixeles que pertenecen a uno y otro marco.

Para determinar la coordenada hasta la que se debe analizar el marco II y no confundirlo con el <marco del número de lista>, se almacena el número de fila de las <marcas de preguntas> en un vector, el cual debe contener más de 48 elementos, los primeros 48 elementos corresponden a la posición de las <marcas de preguntas> del marco II y el elemento 49, indica aproximadamente la fila en la que se ubican las <marcas inferiores del marco del número de lista>.

La coordenada 'x' guardada en el elemento 49 es la que corresponde al primer pixel, con valor menor que el <umbral>, encontrado sobre la recta definida por la <marca lateral derecha> después del elemento 48. Después, el valor del elemento 49 se hace variar 2 pixeles arriba y/o abajo, hasta encontrar 21 <marcas inferiores para determinar el dígito del número de lista>.

Todas las clases de marcas en el formulario tienen un número de error establecido, por si éste llegara a presentarse. El número asociado al error puede localizarse en el código fuente del programa, esto para conocer más a fondo el código si interesa, o bien, para su posible modificación en un futuro.

Una vez que se analizó el marco II del formulario, el programa continua con la identificación del marco I. Para lo cual se sigue el mismo procedimiento que para los tres anteriores.

En este momento, en la imagen del formulario se tiene localizada la posición (coordenada 'x') de cada <marca de pregunta> de los cuatro marcos, con el fin de que puedan distinguirse unas de otras. Cuando se han almacenado estas posiciones en el <vector de preguntas> correspondiente a cada uno de los marcos, se procede al análisis de los <rectángulos de opción> de cada marco.

Para el marco I, el número de fila de las <marcas de pregunta> se almacena de abajo hacia arriba; es decir, la posición de la marca de la pregunta 20 corresponde al primer valor del vector indicado.

Al primer elemento del vector del marco I se le suman 2 unidades, lo que indica 2 pixeles abajo de la <marca de pregunta>, esto es para posicionarse en la fila donde se ubican las <marcas de los rectángulos de opción> de dicho marco. El programa para evaluar los formularios inicia el “recorrido (7)” de derecha a izquierda, almacenando la posición de los <rectángulos de opción> de las preguntas de éste marco. Al finalizar el “recorrido (7)”, se tendrán doce valores almacenados en el vector correspondiente (vector de <marcas de rectángulos de opción> del marco I). Si no es así, existe un error en la imagen que se indica en el archivo correspondiente y el programa evaluará el siguiente formulario.

Si no hubo error en el paso anterior, el programa ubica las <marcas de la clave del examen> del marco I. Al último elemento del vector que guarda la posición de las <marcas de preguntas del marco I> también se le suman 2 unidades, el programa para evaluar los formularios hace un “recorrido (8)” de derecha a izquierda almacenando en el <vector de dígitos de clave> la posición de las <marcas para identificar cada dígito de la clave del examen> (líneas verticales).

Como cada marco tiene diez <marcas de dígitos de clave>, dos para cada cuadro por rellenar, se tendrá opción de clasificar a 125 exámenes de claves diferentes. Los cuadros para la clave están numerados del 0 al 4.

Para ubicar los <rectángulos de opción> del marco II se procede de forma similar que en el marco I, recuérdese que en la localización de cualquier marca siempre se consideran los errores por un punto, por lo que se debe variar 2 pixeles arriba o abajo el valor del píxel que ubica a la marca. El primer elemento del <vector de preguntas> de dicho marco se toma como base y a partir de éste, se hace un “recorrido (9)” de izquierda a derecha para almacenar la coordenada ‘y’ de las <marcas de rectángulos de opción> (líneas verticales). Deben ser doce valores los que se obtengan; si no es así, se escribirá el error en el archivo correspondiente.

Posteriormente se almacena el número de columna de las <marcas de dígitos de la clave> del marco II. Para ello se toma el último valor almacenado en el <vector de preguntas> de dicho marco y se lleva a cabo un “recorrido (10)” de izquierda a derecha para guardar la coordenada ‘y’ de las marcas en el <vector de dígitos de clave>. Al concluir este proceso se tiene identificado el marco número II.

Para el marco III y IV se realiza el procedimiento anterior, se determina el valor de la posición de los <rectángulos de opción> de cada marco (doce valores) y las <marcas de dígitos de clave> del examen (diez valores) y se guarda en su vector correspondiente.

Para concluir con el análisis del formulario es indispensable mencionar que cada uno de los marcos debe tener tres vectores almacenados con tres variables diferentes; un vector para la posición de las preguntas (vector de preguntas), otro para las columnas de los <rectángulos de opción> y el último para ubicar la clave del examen. Si alguno de los tres vectores falta en alguno de los cuatro marcos no se podrá evaluar el formulario.

Cuando se tienen todos los vectores necesarios para evaluar los 4 marcos del formulario se procede con la ubicación del <marco del número de lista>. Se toma como referencia el

elemento 49 del <vector de preguntas> del marco II, debido a que es el más próximo al <marco del número de lista>. A este elemento se le restan 2 unidades y a partir de este valor se “recorre” de izquierda a derecha para guardar el número de columna de las <marcas inferiores que determinan el dígito del número de lista> en un vector (<vector de marcas inferiores del numero de lista>), este vector deberá tener 21 elementos, veinte valores son debido a los dígitos (0-9) y el último es para determinar el fin del marco; sino es así, se le suman 2 unidades al elemento 49 del <vector de preguntas> del marco II y se repite el procedimiento anterior. Si aún así no se obtienen los 21 valores, se descartará el formulario y se describirá el error en el archivo correspondiente.

Para localizar las marcas conectadas al lado derecho del marco del número de lista (<marcas del nivel de significancia para los dígitos del número de lista>), el programa para evaluar se “desplaza” de abajo hacia arriba, con el fin de guardar el número de fila de cada marca como elemento del <vector de marcas del nivel de significancia del dígito del número de lista>. Así se tienen dos vectores para el número de lista de cada formulario, uno para las <marcas inferiores que determinan el dígito del número de lista> (coordenada ‘y’) y el otro para las <marcas del nivel de significancia> (coordenada ‘x’).

Al ubicar todas las marcas en la imagen del formulario digitalizado, se debe considerar además del ruido por puntos, el ruido debido al explorador. Es decir, algunas veces una línea del formulario ocupa en la imagen un píxel como grosor; pero algunas otras ocupa dos o más, aunque se observó que nunca excedía de cuatro píxeles. Por lo tanto, cuando se almacena la posición de las marcas de cada formulario el programa toma en cuenta estas alteraciones y corrige el problema si es posible. Si se encuentra con un punto que no es parte del formulario, se cambia de coordenada (‘x’ o ‘y’ según sea el caso) de búsqueda, si se trata de ubicar la posición de una marca vertical se cambiará la coordenada ‘y’ y si es una marca horizontal se cambiará la coordenada ‘x’. Si de esta manera no se soluciona el problema el programa indicará en el archivo de errores el número y tipo de error suscitado. Es importante señalar que el error por un punto no puede ser corregido si ocupa más de 4 píxeles en la imagen ya sea horizontal o verticalmente.

Cada marco en la imagen del formulario se identifica como una matriz de dimensiones: $m \times n$ y de cada matriz se obtienen copias de pequeñas submatrices que representan a los <rectángulos de opción> (matrices “RO”). Todas las matrices “RO” deberían tener las mismas dimensiones, ya que en el formulario así están dibujadas; pero al digitalizar el formulario se pierden características importantes principalmente cuando el explorador deforma ya sea a las líneas que delimitan a los <rectángulos de opción> o bien a las <marcas de pregunta> de cada marco, con lo que se explica el porqué las matrices “RO” no necesariamente tienen las mismas dimensiones. Es por ello también que se deben localizar y almacenar las <marcas de preguntas> y las <marcas de los rectángulos de opción> para determinar las dimensiones de estas matrices.

Cuando se tienen almacenadas todas las matrices “RO” (cada una de ellas en una variable matricial diferente), se promedian todos sus valores y este dato representa el valor de luminancia promedio que hay en el <rectángulo de opción>, lo que indica si el aspirante sombreó o no esa opción. Si el promedio se aproxima a cero significa que el <rectángulo de opción> en el formulario está sombreado y si se aproxima a 200 no lo está. Recuérdese que

cero es la representación del color negro en la imagen y 255 es del color blanco. Cuando el <rectángulo de opción> no está sombreado, el valor promedio no se aproxima a 255 debido a que en la parte central del <rectángulo de opción> existe la letra que lo identifica (de la 'a' a la 'f').

Por lo anterior, se debe determinar un umbral (valor) que indique si el <rectángulo de opción> está o no sombreado. Después de varios experimentos con imágenes de formularios digitalizados, se determinó que este valor debía ser 90, ya que se observó que si un <rectángulo de opción> está sombreado, el nivel de gris promedio es de 0 a 15 como máximo y cuando no está sombreado el valor se aproxima a 200. Esto sólo resulta si el formulario se contestó con las especificaciones dadas (correcto relleno del <rectángulo de opción>, no manchar si se borra, etc.). Si un aspirante sólo cruzó el <rectángulo de opción> en una o varias preguntas, el promedio obtenido se aproximó a 90. También se observó que al digitalizar un formulario, si el aspirante borró su respuesta de algún <rectángulo de opción> el promedio se aproximaba a 115. Aunque, posteriormente a estas pruebas se acordó que el aspirante debía utilizar plumón negro, y no lápiz para rellenar los <rectángulos de opción> se dejó que el valor del umbral fuese 90.

Para identificar la clave de cada examen y el número de lista del aspirante se procede de la misma forma; se compara el promedio de la matriz "RO" con el del umbral especificado y se decide si está o no sombreado.

Como se mencionó anteriormente, con los valores promedio de la matriz de un <rectángulo de opción> se determina si está o no sombreado. Los resultados de esta "toma de decisiones" se almacenan en un vector con elementos binarios (20 vectores por marco correspondientes a las 20 preguntas de cada examen por evaluar). Un 1 para el <rectángulo de opción> sombreado y cero para el que no lo está. Así, cada vector contiene 6 elementos, cada uno corresponde a un <rectángulo de opción> de la pregunta, el elemento número uno del vector corresponde al <rectángulo de opción> 'a', el número dos al 'b', y así sucesivamente; si todos son cero significa que no hubo respuesta a esa pregunta, si hay dos unos significa que están sombreados dos rectángulos en la misma pregunta, etc. De manera similar se procede para identificar la clave del examen; pero los tres vectores aquí obtenidos se acomodan en una matriz de dimensión: 3x5. En esta matriz no puede haber dos unos en un mismo "renglón", si los hay, no se evaluará ese formulario. Este error no es del programa para evaluar, sino del aspirante por no seguir las indicaciones, ya que en el formulario cada dígito de la clave del examen se debe sombrear en renglones diferentes (nivel de significancia), no en el mismo. Si este error se presentara, se descartará dicho formulario y el programa continuará la evaluación con los siguientes formularios después de escribir el tipo de error en el archivo correspondiente.

Cada marco tiene entonces una matriz (3x5) que representa a la clave del examen y 20 vectores de dimensión 1x6 correspondientes a las 20 preguntas. Como al ejecutar el programa para evaluar los formularios se pidió al operador introducir la clave y las respuestas correctas de los exámenes, se da por hecho que existen variables con las respuestas y las claves de exámenes en la memoria de la computadora.

Para calificar cada examen, primero se comparan las claves proporcionadas por el operador en el formulario virtual con las del aspirante. Como el aspirante tiene opción de escoger dos exámenes de siete más los dos obligatorios y además tiene opción de responder el examen en el marco que desee, las claves de los exámenes indicadas en cada marco del formulario deberán coincidir con las almacenadas en la base de datos; si no es así, el examen se anulará (sólo este examen, no todo el formulario) y aparecerá en el archivo “*calificaciones.txt*” como clave del examen: 000, lo que indica que no se respondió nada en ese marco (aunque sí lo hizo; pero la clave escrita no existe en la base de datos) y como calificación: 0 y el programa continuará con la evaluación del siguiente marco (examen).

Una vez que se haya comparado la clave del examen y no exista ningún error, se procede a comparar las respuestas correctas (formulario virtual) con las del aspirante (formulario digitalizado). Esta comparación se realiza con los dos vectores: si son iguales, la respuesta es correcta, si los vectores son diferentes, la respuesta es incorrecta. Cada uno de los vectores corresponden a cada una de las preguntas del examen por evaluar. ***La respuesta se considera correcta siempre y cuando los dos vectores sean iguales y exista al menos un elemento diferente de cero ('1') en el <vector de respuestas correctas>***; esto se debe al número de preguntas variable en cada examen. Aunque se pueden introducir veinte respuestas en el formulario virtual para cada examen, no necesariamente se deben tener las veinte respuestas, se pueden tener menos. Si no se proporciona respuesta en alguna pregunta del examen, el <vector de respuestas correctas> contendrá sólo valores de cero. Estos datos los completará el operador del sistema para cada uno de los 9 ó 10 exámenes por separado.

En el menú del programa también se pregunta acerca del valor que tendrán las respuestas correctas e incorrectas. Se pueden escoger valores positivos para las respuestas correctas y valores negativos para las respuestas incorrectas como es el caso, o bien 1 para las correctas y 0 para las incorrectas. Esta tarea se deja al operador al momento de ejecutar el programa.

Para obtener la calificación de cada examen, los valores de las respuestas se suman, las respuestas correctas tienen ante omisión un valor positivo (“5”) y las respuestas incorrectas un valor negativo (“-1”) ante la omisión. Finalmente, estos resultados se almacenan en un archivo de texto llamado: “*calificaciones.txt*”. Este archivo almacena el número de lista, y frente a él, las claves de los cuatro exámenes que presentó, todas ellas en fila y con un espacio tabulador y enseguida la calificación obtenida en cada examen, en el mismo orden en el que aparecen las claves; si no contestó alguno de los cuatro exámenes o si no proporcionó clave o la proporcionó mal, aparecerá la clave 000 y como calificación 0.

Cuando se evalúan los formularios, el programa también almacena las respuestas binarias en el otro archivo de texto generado: “*estadísticas.txt*”, esto con el fin de contabilizar la estadística acerca de ¿Cuántos aspirantes respondieron correctamente la pregunta ‘x’ del examen ‘y’?

La estadística se muestra como una gráfica. En el título de la gráfica aparece el nombre del examen, cuántos aspirantes presentaron el examen y abajo se grafica el número de pregunta contra el número de personas que respondieron correctamente, lo que da una idea del nivel de conocimientos del grupo de aspirantes y la dificultad de las preguntas de cada examen.

Como resultado final se obtiene la calificación de todos los aspirantes en un archivo de texto, el cual puede guardarse en la base de datos por medio del programa para evaluar los formularios o bien por medio del programa de la base de datos.

Es importante señalar que los archivos de texto generados durante la evaluación, se conservan en el mismo directorio donde está el programa fuente para evaluar los formularios. Si se desea evaluar nuevamente, se deberán borrar los archivos generados: “*calificaciones.txt*, *estadísticas.txt* y *presentaron.txt*”; de lo contrario, al iniciar la evaluación el programa mostrará una advertencia de que existen los archivos y si no los borra concatenará los datos existentes con los nuevos que se obtendrán.

Si desea mantener las respuestas correctas introducidas en el formulario virtual en un archivo para después utilizarlas para otra evaluación, sólo deberá escribir en el editor de comandos de MATLAB: *save* [nombre del archivo con las respuestas] ⁷ al terminar de introducir las respuestas correctas en el menú, en este archivo se almacenarán todas las variables disponibles en memoria. Si desea cargar el archivo con las respuestas correctas deberá escribir: *load* [nombre del archivo con las respuestas]. El archivo con las respuestas correctas tendrá por extensión “.mat”, este archivo se guardará en el mismo directorio del programa fuente.

⁷ Para obtener más ayuda dirigirse al manual de MATLAB y consultar los comandos: *save* y *load*.

Capítulo V

Generación de estadísticas

Para obtener la estadística de los exámenes presentados: durante la evaluación de los formularios se genera el archivo “*estadísticas.txt*” (Fig. 5.1) el cual contiene las claves de los exámenes presentados y la calificación binaria de los aspirantes (1 = correcto, 0 = incorrecto). El primer valor del renglón es el número de clave del examen presentado, que consta de 3 dígitos y enseguida valores de uno y cero (correcto e incorrecto) que se refieren al número de pregunta de cada examen, separados por un espacio tabulador. El primer valor (1 ó 0) después de la clave del examen corresponde a la pregunta 1, el siguiente valor a la pregunta 2 el siguiente a la 3 y así sucesivamente. El último valor del primer renglón del archivo es el elemento 21 que es el valor binario de la respuesta de la pregunta número 20 del examen (el número de la clave del examen más las 20 preguntas), en el siguiente “renglón” habrá otros 21 elementos correspondientes a las respuestas de otro examen. El total de “renglones” en el archivo será proporcional a 4 veces el número de aspirantes, ya que cada aspirante puede responder 4 exámenes diferentes en el formulario.

100	00100010001000100000
101	11110111011100000100
203	00000000100000000000
210	00000100010101001000
100	00000000010110100000
211	00000000100000000000

Fig. 5.1 Representación de los datos del archivo “*estadísticas.txt*”

Con el programa MATLAB se abre el archivo de texto (“*estadísticas.txt*”) como un archivo secuencial y se guarda en un vector. Como no todos los aspirantes contestan el mismo examen, las claves de los exámenes no estarán ordenadas en ascenso o descenso, sino en el orden en que se respondieron. Por lo que primero, habrá que ordenarlas según su valor numérico y después obtener el total de aciertos para cada pregunta en los diferentes exámenes.

Para obtener el total de respuestas correctas a cada pregunta se deben sumar los elementos de los vectores que corresponden a la misma materia. Si se tiene el archivo de texto como un vector, el primer elemento corresponde a la clave del examen, los siguientes 20 elementos se refieren a las respuestas de ese examen, el elemento 22 es nuevamente otra clave y los elementos que le siguen son sus respectivas respuestas, el elemento 43 es otra clave y así sucesivamente.

Para contabilizar el total de respuestas correctas a cada pregunta de cada examen, se clasifican los resultados según la clave. Primero, se ubica la posición de todos los elementos que tengan la misma clave, una vez identificada la posición se suman los elementos que corresponden a la misma clave, el elemento uno del primer subvector con el uno del siguiente subvector, el dos con el dos, etc., al finalizar la suma, se tendrá un vector de 21 elementos (la clave y las respuestas de las 20 preguntas del examen) que almacena el total de respuestas correctas a cada pregunta de todos los exámenes presentados con la misma clave. Para obtener el total de respuestas correctas en cada examen se procede de la misma forma, al final se deberán tener como máximo diez vectores debido a que el programa evalúa como máximo 10 áreas de conocimiento diferentes.

Los vectores resultantes se almacenan en una base de datos con el programa para evaluar los formularios. Esta base de datos contiene una tabla llamada <estadísticas>, la cual tiene como campos: <clave_examen> y <respondieron> (Fig. 5.2). En esta tabla se almacena la clave del examen y el total de personas que respondieron correctamente el número de pregunta, cuando en la columna (o campo) <clave_examen> se encuentra el valor de la clave por primera vez repetida “n” veces, el último de estos “n” registros contiene, en la segunda columna (o campo) el número de personas que contestaron correctamente la pregunta “n” del examen con dicha clave. Aquí, “n” siempre será $n \leq 20$. Si no se ocupan los 20 registros, sino $m < 20$, los últimos 20-m se rellenan con ceros en la segunda columna y con la misma clave en la primera columna. Por lo tanto, en el campo <clave_examen> se tendrán 20 registros con el mismo número de clave y en el campo <respondieron> se tendrán 20 registros con 20 elementos cada uno correspondiente a cada una de las 20 preguntas del examen (si el examen tiene las 20 preguntas).

Para realizar la gráfica de cada examen se extraen los datos de la tabla <estadísticas> de la base de datos y se opera con ellos. El sistema para evaluar los formularios tiene una opción para llevar acabo este procedimiento; al llamar a esta opción (*graficaestadist.m*) desde el editor de comandos de MATLAB aparece un cuadro de diálogo que pregunta el número de clave del examen por graficar, si esta clave no existe en la base de datos, aparece otra caja de diálogo que indica el tipo de error, si por el contrario la clave existe, se mostrará la gráfica correspondiente (Fig. 5.3). La gráfica tiene como título el nombre del examen, que lo obtiene de la base de datos al consultar el nombre y la clave del examen, y el número de personas que presentaron el examen.

Tanto la base de datos como los archivos de texto y los formularios digitalizados, se almacenan en el mismo directorio del archivo fuente del programa para la evaluación. El resultado final lo constituyen: la base de datos, los archivos de texto (*calificaciones.txt*, *estadísticas.txt* y *errores.txt*, este último sólo si se presentó algún error) y la gráfica (Fig. 5.4).

El sistema para evaluar los formularios almacena la información obtenida en una base de datos. Esta base de datos es creada con Access, es decir, con Access se definen las tablas, sus campos y formatos. Afortunadamente, una vez creada la base se puede manejar con MATLAB. MATLAB se conecta al controlador de las fuentes de datos ODBC de 32 bits instaladas en el sistema operativo Windows y se implementa la programación con instrucciones SQL (Secuency Query Language). Con estas instrucciones se pueden leer, escribir y consultar los datos de cualquier base de datos que haya sido creada con Access (archivos con extensión “mdb”), estos datos son almacenados como vectores (cada columna o campo corresponde a un vector) tanto para realizar cualquier consulta como para poder escribir en la base de datos. Para obtener más información acerca de cómo conectar MATLAB con el programa Access (base de datos) se puede consultar el manual en la página <http://www.mathworks.com>.

	clave_examen	respondieron
14	100	2
15	100	12
16	100	0
17	100	0
18	100	0
19	100	0
20	100	0
21	101	6
22	101	11
23	101	10
24	101	6
25	101	4
26	101	13
27	101	7
28	101	7
29	101	2
30	101	11
31	101	4
32	101	13
33	101	0
34	101	11
35	101	1

Fig. 5.2 Tabla “estadísticas” de la base de datos. En esta figura se consideró como primer registro visible (por la posición de la barra de desplazamiento) el de la pregunta número 14 del examen con clave 100, examen que contiene sólo 15 preguntas y como se observa en la figura, el campo <respondieron> de los registros 16 al 20 se completan con valores de 0. En el campo <clave_examen> del primer registro visible se tiene la clave número 100 y en el campo <respondieron> de dicho registro contiene el valor 2 que corresponde al número de personas que respondieron correctamente la pregunta numero 14 de ese examen. En el campo <clave_examen> del registro número 21 cambia la clave del examen, el 21 y los siguientes 19 registros corresponderán al examen con clave 101 y el registro 21 en el campo <respondieron> se refiere al número de personas que respondieron correctamente la pregunta uno del examen con clave 101.

Los archivos de texto generados no se borran automáticamente al finalizar la evaluación; si desea evaluar nuevamente otro conjunto de formularios se deberán borrar estos archivos manualmente (ver apéndice).

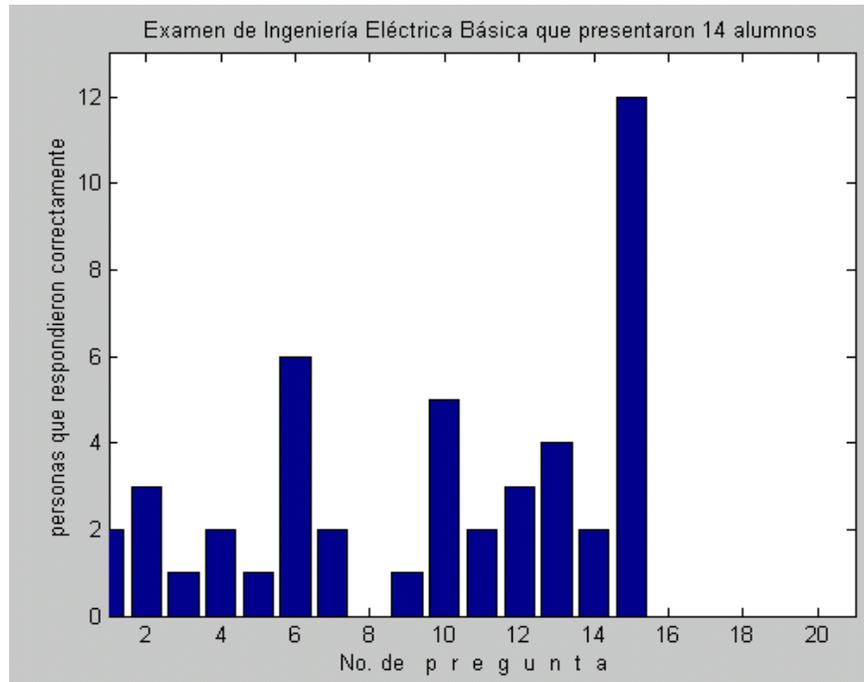


Fig. 5.3 Ejemplo de la gráfica que muestra el programa.

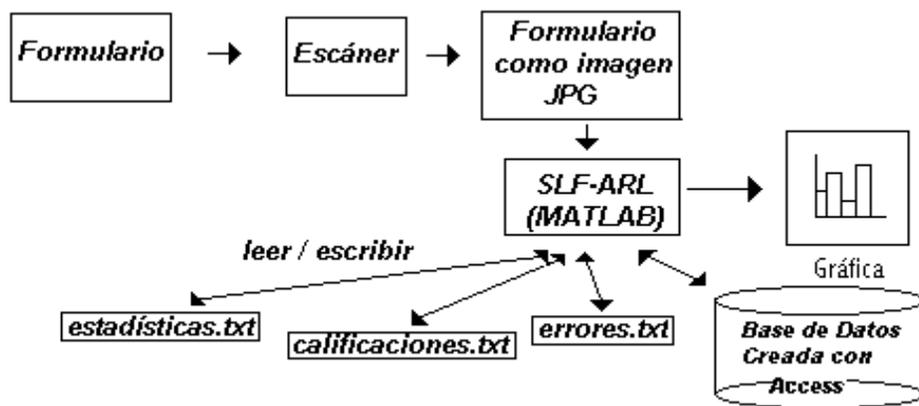


Fig. 5.4 Diagrama de bloques del proceso de evaluación de los exámenes

5.1 Base de datos para el sistema

La base de datos que se ha mencionado, se diseñó con Access para que los datos almacenados en ella no fueran exclusivamente del área de ingeniería.

Las tablas de la base de datos son: <calificación>, <estadísticas>, <nombres>, <materias> y <presentaron> (Fig. 5.5). La tabla <nombres> consta del número de lista y el nombre completo del aspirante. La tabla <estadísticas> tiene sólo dos campos: clave del examen y respondieron; en el campo respondieron se almacena el número de personas que

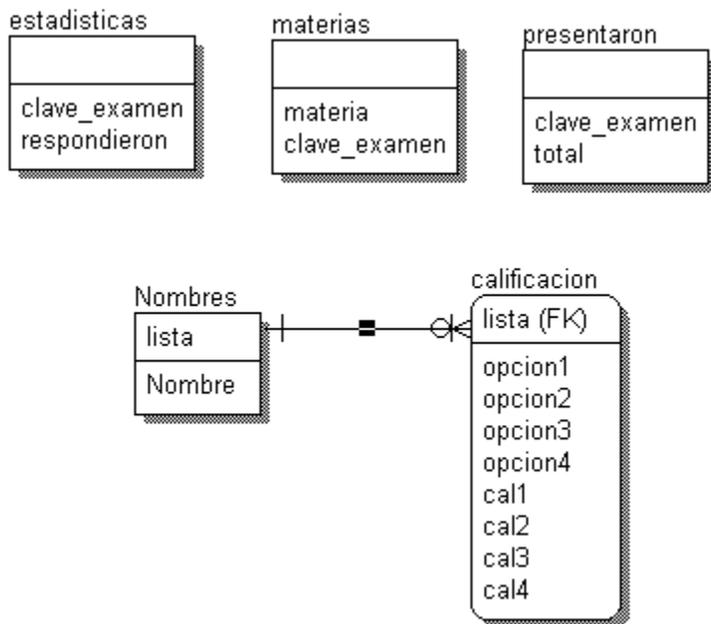


Fig 5.5 Estructura de las tablas de Access

respondieron correctamente el examen con la clave indicada en el campo <clave_examen>, cabe mencionar que aunque esta tabla lleva el mismo nombre que el archivo de texto donde se almacenan las respuestas binarias de los aspirantes (“*estadísticas.txt*”) no contiene los mismos datos y por lo tanto no es lo mismo, ya que la tabla está dentro de una base de datos (Access) y para poder modificar cualquier dato de esta tabla se debe abrir la base de datos y después la tabla <estadísticas>, en cambio el archivo como es de texto se puede modificar con cualquier editor de texto; una base de datos no es lo mismo que un archivo de texto aunque ambos tengan el mismo nombre en Windows se diferencian por la extensión (txt para el archivo de texto y mdb para la base de datos creada con Access). La tabla <calificación> consta de nueve campos, <lista>, <opción1>, <opción2>, <opción3>, <opción4>, <cal1>, <cal2>, <cal3> y <cal4>, el primer campo almacena el número de lista del aspirante, los cuatro campos que le siguen son para guardar la clave de los exámenes que presentó y los últimos cuatro corresponden a la calificación obtenida en cada examen presentado. La tabla <materias> tiene dos campos: clave y nombre de la materia. Finalmente, la tabla <presentaron> también tiene dos campos: la clave del examen y total, éste último campo almacena el total de aspirantes que presentaron el mismo examen, todo esto con fines estadísticos.

El nombre de cada tabla no se escogió como un nombre especial para la evaluación de los aspirantes a maestría, sino que se deja como nombre general para que puedan evaluarse exámenes de diferentes áreas de interés.

Es importante señalar que *antes* de comenzar la evaluación de los formularios, en la tabla <materia>, los campos <clave_examen> y <nombre>, el operador del programa deberá escribir los nombres y claves de los exámenes por evaluar. El nombre y clave del examen en esta tabla sólo se utiliza para la presentación gráfica. Estos datos deben estar completos antes de evaluar los formularios porque el programa que los califica abre la base de datos y obtiene los datos correspondientes para mostrar la gráfica, por lo que si está vacía no mostrará nada. La tabla <nombres>, conserva el nombre completo y el número de lista del aspirante, estos datos no son necesarios para la evaluación, ya que el sistema no los considera para ningún proceso. Los datos en esta tabla se utilizan cuando el operador requiera consultar cuántos y cuales exámenes presentó tal persona, así como la calificación obtenida en cada uno de ellos. La tabla <estadísticas> se completa automáticamente con el sistema, sólo requiere que esté vacía antes de la evaluación, porque si tiene datos almacenados el programa marcará un error. Por esta razón, cuando no se utilicen estos datos se deben respaldar en un archivo (ver apéndice) y borrarlos de la tabla. En la tabla <calificaciones>, se guardan las calificaciones obtenidas de cada examen y no es necesario que esta tabla esté vacía cuando se desee evaluar nuevamente a otra generación, pues estos datos únicamente sirven para consultar los resultados obtenidos en la última evaluación con el programa; pero es recomendable guardarlos con un identificador o etiqueta con Access o respaldarlos con Access también puesto que son tablas de Access. Es importante que el número de lista de los aspirantes no se duplique, si por algún motivo desea agregar datos con números de lista que ya existen, deberá respaldar los datos actuales y borrarlos de la base de datos para que la nueva información no tenga problemas de duplicidad.

5.1.1 Almacenar las respuestas correctas

Las respuestas correctas introducidas en el formulario virtual (Fig. 5.6) se almacenan en diferentes variables cada una de ellas, con el fin de poder comparar sus valores con los obtenidos de los formularios reales (respuestas de los aspirantes).

El formulario virtual consta de opciones tipo radio que simulan a los <rectángulos de opción> y las preguntas de cada marco se etiquetan con un número en la parte lateral izquierda, que indica el número de pregunta. Este formulario virtual es muy similar al formulario diseñado en hoja de papel. Tienen un espacio para marcar la clave del examen y veinte renglones para seleccionar la respuesta (<rectángulos de opción>) de cada una de las 20 preguntas del examen.

El programa para evaluar muestra una pantalla con dos marcos virtuales, así, se tienen cinco pantallas en total para los nueve o diez exámenes por evaluar.

Las respuestas correctas se almacenan en diferentes variables, cada opción de radio es identificada con una variable diferente, como son seis <rectángulos de opción> por

pregunta para cada marco se tendrán 120 variables, por lo tanto, se debe construir el nombre para reconocer cada variable que identifica a cada <rectángulo de opción> incluyendo No. de pregunta y No. de examen. Por ejemplo, para diferenciar cada <rectángulo de opción> de los demás, se consideró la siguiente variable: E1pregunta1(1) (Fig. 5.7a), donde ‘E1’ corresponde al examen, en este caso el examen número 1, como se tienen 9 ó 10 exámenes por evaluar, se tendrán las siglas E1,E2,E3 ... E9 y/ó E10, en esta parte del nombre de la variable, ‘pregunta1’ es la parte del nombre de la variable que indica el número de pregunta (20 en cada marco) y el número que está dentro del paréntesis indica el número ordinal de los <rectángulos de opción> de cada pregunta, 1 para el <rectángulo de opción> identificado con la letra ‘a’, 2 para el <rectángulo de opción> con la letra ‘b’ y así sucesivamente hasta terminar con el número 6 (<rectángulo de opción> con la letra ‘f’).

Como se puede observar para cada marco en el formulario virtual se tienen 20 arreglos con valores binarios (Fig. 5.7b), cada arreglo con seis elementos por pregunta (<rectángulos de opción> de la ‘a’ a la ‘f’). La razón para no usar arreglos con varios subíndices sino definir el nombre de tantas variables, es que ***MATLAB reconoce si un “radio” fue seleccionado o no por medio del valor binario de una variable escalar***, por lo tanto se debe considerar una variable por “radio”, aunque posteriormente se unan seis variables en una sola variable vectorial, el valor de cada una de las variables escalares originales define un elemento en el vector, y al final se tendrá un vector de dimensiones: 1x6.

Clave del examen

0 1 2 3 4

0 1 2 3 4

0 1 2 3 4

1	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
2	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
3	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
4	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
5	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
6	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
7	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
8	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
9	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
10	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
11	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
12	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
13	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
14	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
15	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
16	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
17	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
18	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
19	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
20	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F

Fig. 5.6 Formulario virtual para proporcionar las respuestas correctas del examen

Cabe señalar que todas estas variables vectoriales pueden ser reemplazadas directamente con un sólo vector para todo el proceso de evaluación y no con los que se tiene actualmente; por ejemplo, un sólo vector puede almacenar toda la información de un formulario con un ciclo “do-while”; pero no se dió el caso, debido a que si se almacenara en un vector tantos datos (1 x 480) por formulario, alguna modificación del código del programa no sería muy sencilla, el vector tendría: 6 elementos por pregunta (<rectángulos de opción>) y cada marco tiene veinte preguntas, por lo tanto, para un sólo marco se tendría un vector con 120 elementos y como son 4 marcos resultan los 480 elementos. Más aún, si se quisiera guardar en un sólo vector la información de todas las respuestas correctas (10 exámenes que es el máximo número de exámenes que puede evaluar el programa) se tendría un total de 1200 elementos que almacenaría un solo vector, lo cual dificulta su estudio.

Sea: $Clave_del_examen\#i$, la matriz que contiene la representación de la clave del examen virtual. El subíndice i indica el número de examen: $0 < i \leq 10; i \in \mathbb{Z}$.

$$Clave_del_examen2 = \begin{matrix} & 0 & 1 & 2 & 3 & 4 \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} & \leftarrow & \text{El examen 2 tiene clave 100.} & * \end{matrix}$$

Y sea $examen\#uClave$ la matriz que almacena la representación matricial de la clave del examen del aspirante; $0 < u \leq 4$; $u \in \mathbb{Z}$.

$$examen3Clave = \begin{matrix} \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} & \leftarrow & \text{El marco 3 del formulario tiene clave 100.} & ** \end{matrix}$$

Sea: $E3pregunta12(j)$, la variable binaria que indica la respuesta correcta de la pregunta 12 del examen 3. El subíndice j indica qué rectángulo de opción es el correcto: $1 \leq j \leq 6$ *

$E3pregunta12(3) = 1$; Para la pregunta 12 del examen 3, la respuesta correcta es el rectángulo de opción con la letra 'c'.

Finalmente la variable vectorial $E\#ipregunta_correcta\#k$ almacena las 6 variables binarias con las respuestas dadas a los radios de la pregunta k del examen i . $1 \leq k \leq 20$; $i \in \mathbb{Z}$. Si el examen tiene menos de 20 preguntas los últimos vectores serán de sólo ceros.

$E3pregunta_correcta12 = [0 \ 0 \ 1 \ 0 \ 0 \ 0]$ \leftarrow vector de <rectángulos de opción> "correctos" de la pregunta 12 examen número 3.

Ejemplo de evaluación dadas las variables anteriores :

Sea Z la matriz de respuestas reales (formulario digitalizado) con dimensiones: $m \times n$ (20×6). Si se comparan las variables vectoriales $E3pregunta_correcta12$ (1×6) y $Z(12,1:6)$; $Z(12,1:6)$ es el renglón 12 de la matriz Z y el segundo índice es un rango que corresponde a las 6 columnas de la matriz (vector).

Una vez que se comparó la clave del examen del formulario (matriz) con las claves dadas en el formulario virtual (matrices) y coincidió, se evaluarán las preguntas del examen correspondiente de la siguiente manera: (estos valores son de ejemplo)

Si la respuesta es correcta se tendrá un valor de 5, si es incorrecta -1 y si respondió 0, 2 ó más <rectángulos de opción> en la misma pregunta tendrá valor 0.

Nota: se omiten detalles de la lógica de la programación.

Posibles respuestas del aspirante

calificación de la pregunta 12

- a) $Z(12,1:6) = [0\ 1\ 0\ 0\ 0\ 0]$ -1
- b) $Z(12,1:6) = [1\ 0\ 1\ 0\ 0\ 0]$ 0

- c) $Z(12,1:6) = [0\ 0\ 1\ 0\ 0\ 0]$ 5
- d) $Z(12,1:6) = [0\ 0\ 0\ 0\ 0\ 0]$ 0

* Los valores se definen una vez llenados todos los radios.

** Los valores se definen cada vez que se lee un formulario

Fig. 5.7a Variables con sus posibles valores para las respuestas correctas, incorrectas y nulas.

Clave del examen

0
 1
 2
 3
 4

0
 1
 2
 3
 4

0
 1
 2
 3
 4

1	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
2	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
3	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
4	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
5	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
6	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
7	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
8	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
9	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
10	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
11	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
12	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
13	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
14	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
15	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
16	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
17	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
18	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
19	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F
20	<input type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E	<input type="radio"/> F

} Los valores binarios se depositan en la matriz Clave_del_examen#i

} Los valores binarios se depositan en la variable E#ipregunta1

⋮

⋮

} Los valores binarios se depositan en la variable E#ipregunta7

⋮

⋮

} Los valores binarios se depositan en la variable E#ipregunta20

Fig. 5.7b '#i' es el valor numérico de la variable 'i', el cual forma, dinámicamente, parte del nombre de las variables referenciales. <<Clave_del_examen#i>> es una matriz de dimensión 3x5; i=1,2,3,..., [9 ó 10]

Capítulo VI

Resultados

Una vez implantado este sistema para evaluar formularios se hicieron las primeras pruebas para conocer su desempeño. Para lograr los resultados esperados con este sistema se consideraron aspectos fundamentales tales como: el tipo de lápiz o tinta utilizado al rellenar los <rectángulos de opción> en los formularios, las manchas producidas al borrar los <rectángulos de opción> cuando uno se equivoca, el correcto e incorrecto rellenado de los <rectángulos de opción>, y finalmente, todas las posibles opciones al digitalizar la imagen: giros, desplazamientos sobre la cama del explorador, etc.

Se hicieron 4 pruebas con diferente número de formularios, usando siempre formularios impresos originales (nunca fotocopias).

La primera prueba se hizo con 20 estudiantes de la Facultad de Ingeniería, y resultó que 16 de los formularios digitalizados fueron evaluados con éxito y 4 presentaron error; los detalles de error para estos formularios fueron: al digitalizar los formularios, el explorador no detectó algunas líneas de los <rectángulos de opción> y en algunos otros no se reconocieron las líneas que definen la clave del examen, por lo que se prosiguió a *aumentar el grosor de dichas líneas* en el formulario original (hoja tamaño carta) con el programa MS Word para corregir este detalle.

La segunda prueba fue para 30 personas de diferente nivel cultural, ya con las correcciones indicadas los problemas anteriores se solucionaron, pero se presentaron 3 formularios con error, estos errores ahora eran debidos a manchas introducidas en el centro del formulario *al digitalizar*, por lo que se tuvieron que *volver a digitalizar* los formularios. Al digitalizarlos nuevamente y evaluarlos con el programa, sí se evaluaron correctamente los tres formularios.

La tercera prueba fue en un seminario para 40 alumnos de maestría, en el cual se hizo una evaluación a las ponentes del seminario. Aquí se presentaron 7 errores porque los encuestados rayaron el formulario, no lo sombrearon correctamente, alteraron el formulario y algunos otros errores a consecuencia de fallas del explorador. Una vez corregidos estos problemas se evaluaron correctamente los formularios. La corrección consistió en editar la imagen digitalizada con el procesador de imágenes Igrafx, y para cada formulario fue diferente, ya que es función del error indicado en el archivo "errores.txt" del evaluador de formularios. Cuando el error se debió al explorador, sólo se borraron de la imagen JPG los puntos negros que no están definidos en el formulario original, en los casos de rayado y marcas indebidas, se borraron las líneas alteradas y posteriormente se completaron en forma adecuada. Todas las imágenes corregidas se guardaron nuevamente con formato JPG y factor de compresión 1.

En el caso de que el error haya sido por rotar el formulario al momento de digitalizar, es necesario volver a digitalizar el formulario. Si el operador corrigió algún problema con el

editor de imágenes y al momento de evaluarlo con el sistema se vuelve a presentar el error, es recomendable volver a digitalizar el formulario.

Cabe señalar que la mayoría de los formularios no evaluados se debieron al mal uso del formulario por parte de los estudiantes al rellenar las opciones.

La cuarta prueba fue para 60 alumnos de la Facultad de Ingeniería y se presentaron 4 errores consecuencia del explorador, los cuales al *digitalizarlos nuevamente* se evaluaron con éxito.

Es importante señalar que los errores presentados durante la evaluación de los formularios pueden ser cometidos por el operador al digitalizar la imagen, por el explorador, o bien, por la persona que contesta el formulario. Lo importante es que el sistema detecta estos errores y los escribe y tipifica en el archivo correspondiente, advirtiéndolo al operador para su corrección.

Todas las evaluaciones que se ejecutaron sin errores detectados por el programa fueron confiables, lo que se puede afirmar porque se verificó manualmente cada una de las respuestas, para lo cual se consideraron los errores de los encuestados, tales como: <rectángulos de opción> mal sombreados, <rectángulos de opción> mal borrados, dos o más respuestas en una pregunta, etc.

Para evitar errores por parte de los aspirantes al llenar los formularios, se diseñó un manual que indica cómo se debe sombrear los <rectángulos de opción> de cada pregunta, y además señala los errores que el sistema detecta por alteraciones al formulario.

Para operar el programa llamado «evaluación de formularios», se diseñó otro manual, el cual indica el proceso para la evaluación, los errores que se pueden corregir, la posición para digitalizar los formularios, el diseño de la base de datos y la información necesaria antes de evaluar.

Para finalizar, la calificación de cada formulario con este sistema toma aproximadamente un segundo con las características de la computadora antes señaladas; se tiene en cuenta que la imagen ya ha sido almacenada en el formato indicado. El tiempo para todo el proceso (digitalizar el formulario y evaluarlo) es aproximadamente de 1 minuto. Según el número de formularios por evaluar será el tiempo para la evaluación; aunque el tiempo más considerable es el requerido para la digitalización de la imagen.

Al concluir la evaluación y corrección de todos los formularios, la información queda almacenada tanto en la base de datos como en los archivos de texto, y además, si uno lo desea, se puede obtener la gráfica correspondiente para cada examen.

Capítulo VII

Conclusiones

Se concluye que el sistema lector de formularios será de gran utilidad para la evaluación de los aspirantes a maestría y doctorado, el tiempo para la entrega de resultados se reducirá considerablemente, y la evaluación tendrá un alto grado de confiabilidad, además, la calificación obtenida por cada aspirante se conservará en la base de datos. Una sola persona podrá evaluar los formularios y no requerirá de amplios conocimientos en el área, y además, se emplean dispositivos que se pueden adquirir fácilmente en el mercado y a un costo bajo.

Aunque existen dispositivos comerciales para hacer estas evaluaciones, se da una alternativa que resulta más económica porque no requiere de mantenimiento, sólo de un explorador, una impresora y una computadora; herramientas cotidianas en cualquier oficina. Los formularios se pueden imprimir con impresora laser directamente en hojas de papel bond tamaño carta, o bien en una imprenta. No se requiere de mucho para obtener buenos resultados.

Las *limitantes* que presenta el sistema son: los formularios por evaluar son únicos, es decir, este sistema se ajusta al diseño de la hoja de respuestas, no funciona para otro formulario con diferente diseño, ocupa cierto espacio de memoria en el disco duro, sobre todo las imágenes exploradas (243 Kb), aunque después de su evaluación pueden eliminarse. Se necesitan los programas MATLAB, iGrafx, MS Access, MS Word y el del explorador, ya que sin ellos no se llevará a cabo el proceso de evaluación. Si son demasiados formularios por evaluar (más de 100) resulta tedioso explorarlos, aunque se puede dividir el trabajo al explorar unos formularios un día y otros al siguiente día y así sucesivamente, hasta tener todos almacenados y proceder luego a su evaluación. Si el explorador introduce mucho ruido o bien se producen demasiados errores no será práctico volver a explorar todos los formularios, mucho menos corregir cada uno de ellos con un editor de imágenes.

La principal limitante es cuando se requiere evaluar un gran número de personas, ya que el tedio de explorar es proporcional al número de formularios, pero afortunadamente en el mercado existe un explorador con alimentador de documentos (*ADF: Automatic Document Feeder*), lo cual automatizaría aún más el proceso de evaluación. En el mercado hay exploradores con alimentador de 20 a 30 documentos, con esto se tendría un proceso menos tedioso y con menos errores por parte del operador, aunque aún son caros estos dispositivos (\$499 dls.), y difíciles de encontrar, se espera que con el tiempo sean más accesibles y de mayor capacidad.

APÉNDICE

MANUAL DEL SISTEMA PARA EVALUAR LAS HOJAS DE RESPUESTAS DE LOS ASPIRANTES A MAESTRÍA Y DOCTORADO. (SLF - ARL)

Objetivos:

Este manual pretende explicar los pasos que se deben seguir para evaluar los formularios que contienen las respuestas de los exámenes de aspirantes del nivel maestría y doctorado utilizando el sistema desarrollado para *Lectura de Formularios*.

Pretende también orientar sobre las causas posibles de los errores que se presenten durante la evaluación.

Es importante hacer notar que los programas que se utilizan son: iGrafx™¹ (para procesar la imagen), MATLAB™² (Ver. 6.0), MS-Access (para crear la base de datos), y MS-Word para abrir e imprimir el formulario; además, requiere que el digitalizador de imágenes (“*scanner*”) esté instalado en la PC.

Nota: Si alguno de estos productos de “*software*” falta, no se podrá llevar a cabo la evaluación.

Procedimiento.

1. Definir la estructura de la base de datos en donde se guardarán las calificaciones, los datos obtenidos de los formularios, además de los datos personales de los aspirantes.
- 2.- Digitalizar los formularios y guardar la lectura resultante en formato JPG.
- 3.- Ejecutar el sistema para evaluar los formularios.

1-i. Si es la primera vez que se evalúa:

1-i-a. Estructuración de la base de datos (sistema de tablas)

Si es la primera vez que se va a estructurar la base de datos, se deberán abrir las fuentes de datos ODBC, para permitir la conexión entre MATLAB y MS-Access.

Incite con el botón izquierdo del ratón en el menú inicio de la barra de Windows.

¹ Fabricado por Micrografx, Inc.

² Mathworks, Inc.

Posiciónese en *configuración* y presione en la opción .

Ahora, oprima dos veces en *Fuente de datos ODBC (32 bits)*.

En la pestaña de la ventana **DNS de usuario** incite , seleccione *Microsoft Access driver (*.mdb)* y oprima el botón .

Entonces, aparecerá una ventana para la configuración de ODBC Microsoft Access. Escriba "Prueba2Tesis" en el campo para el '*Nombre del origen de datos*', deje vacío el campo para la '*descripción*' e incite que se encuentra inmediatamente abajo.

En la nueva ventana elija la ruta C:\oscar\pmm y escriba "Prueba2Tesis" en el campo para el '*Nombre de la base de datos*', y oprima . Aparecerá un mensaje que describe el resultado de la inicialización de la base (correcta o incorrecta). Cierre ODBC mediante tres aceptaciones consecutivas.

La base de datos ya está lista en el directorio debido:

C:\oscar\pmm\Prueba2Tesis.mdb.

Contenido de la Base de Datos

NOTA: Los nombres de las tablas y los campos los deberá escribir sin acento.

En la base de datos creada, se deben delinear 5 tablas, la primera guardará la calificación obtenida por el aspirante en el examen correspondiente, la segunda almacena datos estadísticos, la tercera almacena el número de alumnos que presentaron cada uno de los exámenes la siguiente tabla guarda las claves de las materias con su respectivo nombre y la última tabla contiene el número de lista y nombre de los aspirante.

1. Ejecute la aplicación MsAccess: → → .
2. Abra la base de datos que inicializó en el punto anterior seleccionando el archivo C:\oscar\pmm\Prueba2Tesis.mdb y presione el botón .
3. Presione 2 veces el botón izquierdo del ratón en la opción *crear una tabla en vista Diseño* dentro del mismo menú.
4. Aparecerá una ventana correspondiente a la tabla, debe escribir en '*Nombre del campo*': lista y en el campo '*Tipo de datos*', seleccione *numérico*, escriba en el siguiente renglón: opcion1 para el '*Nombre del campo*' y en '*Tipo de datos*': *numérico*, en el siguiente renglón escriba: opcion2 y es de tipo '*numérico*', como siguiente campo escriba: opcion3, como tipo '*numérico*', nuevamente para el nuevo campo opcion4 de tipo '*numérico*', los siguientes 4 campos son para las calificaciones obtenidas en los cuatro exámenes anteriores, el campo

correspondiente a la calificación de la primera clave de examen es: cal1, el segundo cal2, cal3 y cal4 para los dos últimos todos de tipo 'numérico'.

5. Guarde la tabla presionando en el menú de MS-Access el botón → opción *Guardar* aparece una ventana con un espacio para escribir el nombre de la tabla, escriba: **calificacion** y presione el botón . Aparece un mensaje que indica que ¡ No hay ninguna clave principal definida ! y que si ¿ desea crear una ? presione el botón .

6. Debe crear la segunda tabla repitiendo los pasos anteriores a partir del número 3, los nombres de los campos son: clave_examen (numérico) y respondieron (numérico), el nombre de la tabla es: **estadística**.

7. La tercera tabla se crea de forma similar, los campos son: clave_examen y total, ambos de tipo numérico, el nombre de la tabla es: **presentaron**.

8. La cuarta tabla contiene los campos: materia (texto) y clave_examen de tipo numérico.

9. Por último deberá delinear la tabla **Nombres**, la cual almacenará el nombre del aspirante y su número de lista. Los campos de esta tabla son: lista (numérico) y Nombre (texto).

En total habrá 5 tablas en la base de datos, si alguna de ellas falta no se evaluará correctamente.

Llene la tabla: **materias**, con los datos correspondientes al nombre de la materia y su clave. Se deben tener 9 ó 10 registros en esta tabla, ya que sólo se pueden evaluar como máximo 10 diferentes exámenes y como mínimo 9, teniendo en cuenta que si no se ha completado esta tabla no se evaluará correctamente.

Llene la tabla: **Nombres**, la cual consta de los datos del aspirante (Nombre completo y su número de lista). Esta tabla es opcional, pero se recomienda llenar antes de la evaluación ya que para hacer cualquier consulta sobre la calificación, es más fácil obtener resultados si se tienen estos datos.

1-ii. Si no es la primera vez que se evalúa:

Como la base de datos ya existe, sólo se requiere que las tablas **estadísticas**, **calificacion** y **presentaron** estén vacías, para lo cual se deberá respaldar la información actual en otro directorio o bien en disco flexible, si así lo desea, y borrar los datos contenidos en ellas, como se describe a continuación:

PARA GUARDAR LA BASE ACTUAL. (opcional)

Expande el ícono en la ventana principal de Windows, diríjase al directorio C:\oscar\pmm\ y seleccione el archivo "Prueba2Tesis.mdb" con el botón derecho del ratón, posicione en *enviar a*, seleccione *disco de 3 ½ (A)* y ejecute.

PARA BORRAR LOS DATOS DE LA BASE ACTUAL

Abra la base de datos C:\oscar\pmm\Prueba2Tesis.mdb

Abra la tabla **estadísticas** presionando dos veces en la etiqueta que contiene este nombre. Una vez abierta la tabla en el menú *Edición* seleccione *Todos los registros*, después en el mismo menú *Edición* seleccione *Eliminar*, guarde los cambios y cierre la tabla.

Abra la tabla **calificación**. Incite en el menú *Edición*, seleccione *Todos los registros* y en el mismo menú *Edición* ejecute *Eliminar*.

Aplique los pasos anteriores para borrar los registros de la tabla **presentaron**.

NOTA: Las tablas anteriores deben estar vacías antes de calificar para que se obtengan los resultados esperados.

2. Exploración de los formularios (hojas de respuestas)

Los formularios por explorar son las hojas impresas que contienen las respuestas de los aspirantes. Se recomienda que éstas, se ordenen en forma ascendente según el número de lista. Para que al grabar la imagen en la computadora se guarde con algún nombre seguido del número de lista (sin espacio entre ambos). Haciendo más fácil la ubicación del formulario si se presenta algún error.

2.1 Iniciar la evaluación.

1. Conecte el explorador a la alimentación eléctrica así como a la PC.
2. Coloque el primer formulario en la esquina superior izquierda del cristal del scanner.
3. Presione el botón del menú , seleccione → e incite la opción *Image*.
4. En la aplicación iGrafx presione la opción *File* del menú y seleccione *Acquire* (Shift + F3). "Comenzará la exploración del scanner en otra ventana con su propia aplicación: Hp Precision Scan LTX".
5. En la aplicación HP Precision Scan seleccione *Ajustes* opte por *tipo de salida*, seleccione *dibujo en blanco y negro* y presione el botón .

En el mismo menú *Ajustes* , opte ahora por *Resolución* y presione en la etiqueta *Establecer una resolución personalizada*, escriba con número "75" y presione el botón .

- 6).- Presione el botón .

Nota: Si llega a aparecer un mensaje de error al terminar de enviar la imagen sólo oprimir el botón .

Automáticamente regresará a la aplicación iGrafx seleccione la opción *File* del menú y presione en la opción *Save As*. Escriba la ruta en donde esta la aplicación **SLF – ARL** (“C:\oscar\pmm”).

Debe guardar las imágenes en forma consecutiva, para ello en el campo ‘*nombre del archivo*’ debe colocar el nombre seguido de un número (por facilidad el número de lista), por ejemplo a1, a2, a3, ..., etc. Por eso se recomienda ordenar en forma ascendente los formularios antes de digitalizarlos.

En el siguiente campo *Guardar como*, debe seleccionar *JPEG File Interchange (*.jpg)*. De lo contrario no se evaluarán los formularios.

Finalmente presionar el botón .

Aparecerá una ventana que contiene las opciones para guardar la imagen JPG, tal como factor de compresión entre otros. no altere nada. Presione el botón .

Repetir el procedimiento hasta terminar con todos los formularios.

Nota: Las imágenes deberán guardarse con las mismas iniciales y al final un número consecutivo que las distinga, Si alguna imagen no cumple con esta indicación no se evaluará. Por lo cual se recomienda guardarlas con una letra seguida de un número por ejemplo: a1.jpg, a2.jpg, a3.jpg, a4.jpg, ... ,aN.jpg; donde N es igual al número de formularios.

Cabe señalar que las imágenes deben guardarse dentro del directorio en el cual está la aplicación para evaluarlos. (“C:\oscar\pmm”).

3. Ejecutar la aplicación en MATLAB para la evaluación de los formularios.

1. Para ejecutar MATLAB abra el menú *inicio* de Windows, posicione en la opción *programas* y presione 2 veces en la etiqueta *MATLABR12*.

2. Escriba en la línea de comandos de MATLAB la entrada “*cd C:\oscar\pmm*” y dé el «retorno de carro» para cambiar al directorio donde está la aplicación.

3. Escriba “*borra3*” para eliminar, si están allí, los archivos “errores.txt”, “calificaciones.txt”, “estadísticas.txt”. Si no se encuentran estos archivos, no sucede nada.

4. Para ejecutar la tarea escriba, en la línea de comandos del intérprete MATLAB, la entrada “*calificar*” y presione la tecla de «retorno de carro». Aparecerá un mensaje que presenta la opción para evaluar 9 ó 10 *formularios*.

5. Presione para 9 *formularios*, para 10 *formularios*.

6. Aparece un menú que permite indicarle al programa la clave del examen por evaluar y el valores de las respuestas correctas e incorrectas del examen correspondiente. Debe introducir todas las claves de los exámenes por evaluar así como el valor de sus respuestas correctas e incorrectas, si alguna clave falta el sistema marcará error al momento de la evaluación.

Presione la opción “*Almacenar respuestas correctas, Examen I y II*” y maximice la ventana correspondiente a esta opción, seleccione con el botón izquierdo del ratón los tres dígitos de la clave del examen por evaluar y posteriormente las letras que coinciden con las respuestas correctas, dejando en blanco los círculos (comúnmente llamados *radios*) que no tengan respuestas (en caso de señalar dos círculos del mismo renglón, la pregunta correspondiente será calificada como mal contestada).

Una vez que termine de señalar las respuestas correctas para los dos exámenes seleccionados, presione el botón y presione luego en la barra del menú de Windows el botón para regresar al MENU.

7. Repita el paso anterior hasta llegar al examen 9 y/ó 10.

8. En el menú del sistema presione el botón para comenzar con la evaluación.

9. En el siguiente paso aparece una ventana de dialogo que indica la forma en que se debe ingresar el nombre de cada archivo a evaluar, si presiona “NO” el programa procederá a la evaluación de todos los exámenes sin consultar al operador, pero si contesta “SI”, el operador tendrá que escribir uno a uno el nombre de cada archivo por evaluar (sin la extensión, pero sí con el número de orden de lectura, por ejemplo “a1”). El formulario por evaluar es JPG y no evaluará otro diferente; entonces, se supone que el operador tuvo el cuidado de que las exploraciones de los formularios se guardaron en dicho formato JPG. Al presionar el botón de esta ventana, aparecerá la ventana descrita en el párrafo siguiente.

10. Esta ventana es la final, si contestó “NO” en 9. Tiene dos campos por contestar: como se optó por evaluar automáticamente todos los formularios, el programa pregunta el [*nombre del primer archivo*], y el [*total de exámenes por calificar*]. El nombre del primer archivo lo debe escribir sin el número; por ejemplo “a”, ya que automáticamente el programa le asignará el número 1, y se evaluarán consecutivamente todos los formularios hasta el número que se le dio en el campo [*total de exámenes por calificar*]. Si alguno de los archivos por evaluar no existe o

presenta problemas graves para su evaluación, aparecerá, dentro de un archivo “*errores.txt*”, el *número de error* y el *nombre secuencial del archivo* en el que se suscitó la dificultad.

11. Presione el botón **salir** del menú del programa.

12. Si desea almacenar las respuestas correctas en un archivo para su uso posterior, debe escribir en el editor de comandos de MATLAB: “*save [nombre del archivo]*” y si desea evaluar nuevamente otros formularios y tiene el archivo con las respuestas correctas, debe escribir lo siguiente: “*load [nombre del archivo]*” para cargar las variables con las respuestas correctas.

Para saber si se evaluaron correctamente todos los formularios o si hubo errores, debe consultar el archivo *errores.txt*, el cual se crea automáticamente en el directorio donde están las imágenes (en este ejemplo, C:\oscar\pmm); si este archivo no aparece es porque no hubo error alguno ahora, y se han borrado correctamente las versiones anteriores con *borra3*. Si existe, puede editarlo con el comando “*edit errores.txt*” de MATLAB, o bien con el “*Notepad*” de Windows o con cualquier otro editor de texto, para ver el *número de error* y una descripción breve correspondiente. Si se desea saber más sobre un error particular, con el número del error se puede localizar su descripción completa a través del examen del código del programa fuente “*sum_cuadrado.m*”.

12. Para corregir los errores existen dos formas:

12.1. La más sencilla, explorar nuevamente el formulario.

12.2. Corregir el archivo JPG correspondiente al formulario problemático con el lápiz virtual de un editor de imágenes, tal como iGrafx™.

¡NO CIERRE LA APLICACIÓN DE MATLAB PARA CORREGIR ERRORES!
(Solo minimice la ventana).

Debe borrar el archivo “*errores.txt*” para que al evaluar nuevamente el conjunto de formularios problemáticos, pueda ud. saber si se volvieron a presentar los errores.

¡NO BORRE *errores.txt* EN ESTE MOMENTO CON *borra3*!

Si la imagen problemática lo es porque está girada, no intente corregirla con la opción de rotación del editor de imágenes, porque aunque luzca correcta, en el nivel de pixel estará defectuosa.

Algunos errores no se podrán corregir, por ejemplo: cuando el formulario esté rotado por causas ajenas al operador (por ejemplo, al imprimir el formulario), cuando el formulario esté doblado o arrugado, o en un caso extremo, cuando el lector de imágenes no explore correctamente los formularios.

Los errores que sí se pueden corregir con el editor de imágenes son: cuando en la imagen aparecen manchas y marcas, ya sean resultantes de la exploración misma o causadas por el aspirante, se deben borrar con la aplicación iGrafx o cualquier otro editor de imágenes. Cuando el formulario esté de cabeza se debe colocar en la posición correcta rotándola 180° o volviéndola a explorar.

DEBE VOLVER A EXPLORAR EL FORMULARIO CUANDO:

El operador haya cometido algún error como: no guardar la imagen, no usar la resolución especificada (75 dpi), no guardarla con el nombre apropiado o en el directorio especificado, y poner el formulario inclinado o rotado sobre la cama del lector.

Una vez corregidos los errores anteriores (si los hubo), se deben evaluar nuevamente los formularios con problemas; para esto:

12.3. Ejecute la aplicación “*pregunta*” del sistema escribiendo *pregunta* tal cual en la línea de comandos de MATLAB y dé <<retorno de carro>>.

12.4. Aparecerá la ventana que indica si desea evaluar una por una, debe borrar el “NO” y escribir “SI” en el primer campo, posteriormente presione el botón .

12.5. Escriba el *nombre secuencial* del archivo por evaluar; recuerde que el *nombre secuencial* incluye el número y excluye la extensión, por ejemplo: “a1”. Presione .

12.6. Al evaluar este archivo aparecerá otra ventana que da opción de evaluar otro formulario. Presione “SI” para evaluar otra hoja de respuestas, “NO” para terminar la evaluación.

Al finalizar la evaluación de todos los formularios aparecerán los siguientes archivos:

- i calificaciones.txt*, que contiene el número de lista de los aspirantes, la clave de los exámenes presentados y la calificación obtenida en cada uno de ellos,
- ii estadísticas.txt*, el cual contiene los datos referentes a las estadísticas y,
- iii errores.txt*, si es que se presentaron errores al evaluar.

13. Ejecute la aplicación “*meteaDB*”, escribiendo “*meteaDB*” tal cual en la línea de comandos de MATLAB. Esta aplicación guarda las calificaciones y el número de lista del aspirante en la base de datos denominada “*prueba2tesis*”.

14. Ejecute el programa “*presentaronDB*”. Esta aplicación guarda en la base de datos “*prueba2tesis*” el número de aspirantes que respondieron cada uno de los exámenes (Cuántos aspirantes presentaron el examen 1, cuántos el 2, etc.).

15. Ejecute el programa *leestadistic*. Esta aplicación guarda los datos contenidos en el archivo *estadísticas.txt* en la base de datos *prueba2tesis*. Si este archivo *estadísticas.txt* no existe, MATLAB indicará un error.

16. Ejecute el programa *graficaestadist* para mostrar la gráfica correspondiente al examen indicado. Durante la ejecución le pedirá la clave del examen cuyos datos se contabilizarán. Este programa toma sus datos de entrada directamente de la base *prueba2tesis*.

Bibliografía

Norton, Peter. *Periféricos y accesorios para la IBM-PC, PS/2 y compatibles*. Prentice Hall Hispanoamericana S.A., 1993.

Blatner, David. Fleishman Glenn and Roth Steve. *Real World Scanning and Halftones*. Peachpit Press, 1998.

AWCOK, G.W and Thomas R. *Applied Image Processing*. The Macmillan Press LTD, 1996

González, Rafael. *Digital Image Processing*.

Apuntes de Procesamiento Digital de Imágenes. Facultad de Ingeniería, UNAM.

Aguilar García Claudia Erika. *Evaluación y Modelado e Instrumentación del proceso de admisión al Posgrado de Ingeniería Eléctrica*. Tesis, UNAM Facultad de Ingeniería, 2000, 28,32,34,36,38 pp.

Páginas de internet:

<http://www.dara.es/omr/Present.htm>

<http://www.mathworks.com>

<http://www.w3.org/graphics/JPEG>

<http://ece.pordue.edu/ace/jpeg-tut>

<http://adagio.music.northwestern.edu/>