

UNIVERSIDAD CENTRAL DEL ECUADOR

FACULTAD DE INGENIERÍA, CIENCIAS FÍSICAS

Y MATEMÁTICA

CARRERA DE INGENIERÍA INFORMÁTICA

SISTEMA DE RECONOCIMIENTO ÓPTICO DE CARACTERES, COMO SOLUCIÓN AL PROCESAMIENTO DIGITAL DE LA INFORMACIÓN EN LA FACULTAD DE INGENIERÍA, CIENCIAS FÍSICAS Y MATEMÁTICA ESPECIALIDAD – INFORMÁTICA

TRABAJO DE GRADUACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO INFORMÁTICO

AUTORES:

Betancourt Salazar Silvana Del Rocío

Ponce Proaño Miguel Alejandro

TUTOR:

Ing. Franz Del Pozo

QUITO - ECUADOR

2012

## Manual de usuario

El Sistema Ocr, es una propuesta orientada al proceso de digitalización y reconocimiento óptico de caracteres en imágenes digitales que sustenta el trabajo de digitalización de forma automática. Este aplicativo está basado en formas Windows es decir es una aplicación de escritorio, el objetivo del mismo es procesar archivos de imágenes digitales y exportarlos hacia archivos con múltiples formatos como: word, excel, pdf, txt y xml.

**Objetivo**

El principal objetivo de este manual es facilitar el manejo del sistema Ocr, permitiendo establecer el proceso sobre el manejo de las diferentes opciones y utilidades, en síntesis indica la funcionalidad de cada componente que contiene el sistema.

**Características generales del sistema**

* El sistema permite la ejecución automática en intervalos de tiempo.
* El sistema permite seleccionar los directorios origen y destino para los distintos tipos de archivos.
* El sistema muestra un resumen de la cantidad, porcentaje y tiempo utilizado por lote de procesamiento.
* El sistema maneja tres tipos de estados para los documentos: procesados, no procesados y error en el proceso.
* El sistema muestra una barra de progreso, que indica el porcentaje de procesamiento de documentos y de manera análoga para el proceso de documentos no procesados en primera instancia.
* El sistema maneja un módulo para validar la conexión con el servidor de base de datos.
* El sistema maneja un módulo para la actualización de la bandera Ocr, en la base de datos.

**Alcance del sistema**

El Sistema permite procesar toda la documentación previamente digitalizada con formato tif, mediante subsistemas de procesamiento y reconocimiento Ocr, con módulos para establecer la conexión y actualización de los registros en la base de datos.

**Pantalla principal del sistema**

El Sistema de gestión documental es una aplicación Windows.

Para ingresar al aplicativo ingrese a través de su icono ejecutable en el escritorio de Windows: **Ocr.FacultadIngenieria**

Inmediatamente se presenta la pantalla principal del aplicativo.

Gráfico 61. Pantalla principal del aplicativo



Autores: Silvana Betancourt, Miguel Ponce

**Botón Procesamiento:** Visualizara un conjunto de tabs que se explicara más adelante

Gráfico 62. Botón procesamiento



Autores: Silvana Betancourt, Miguel Ponce

Menú principal del botón de procesamiento, Aquí visualizara varios tabs como:

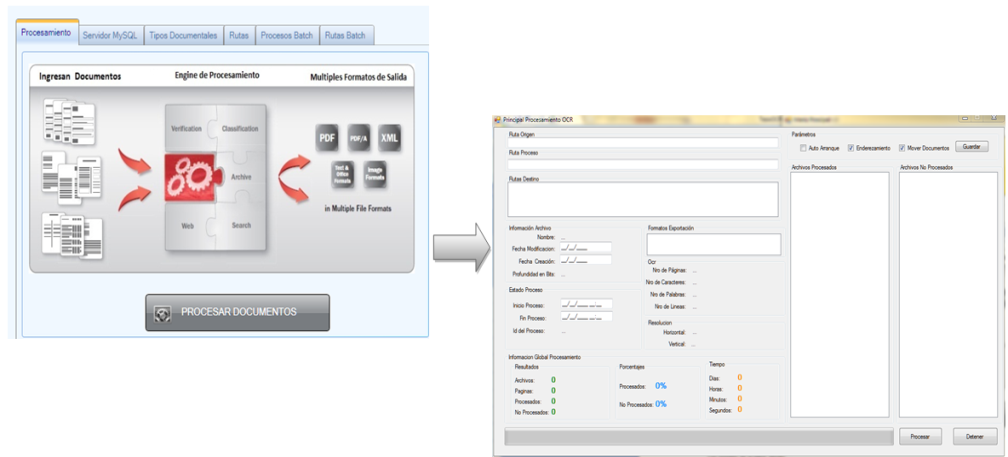
Gráfico 63. Tabs de la pantalla principal



Autores: Silvana Betancourt, Miguel Ponce

**Tab Procesamiento:** ventana utilizada para el procesamiento de documentos, aquí se puede visualizar un resumen de los resultados por lote de documentos procesados, porcentaje de documentos procesados y el tiempo utilizado para el proceso de reconocimiento.

Gráfico 64. Procesamiento



Autores: Silvana Betancourt, Miguel Ponce

Los botones en la parte derecha de la ventana de procesamiento permiten realizar las siguientes acciones:

Gráfico 65. Botones de acción formulario principal

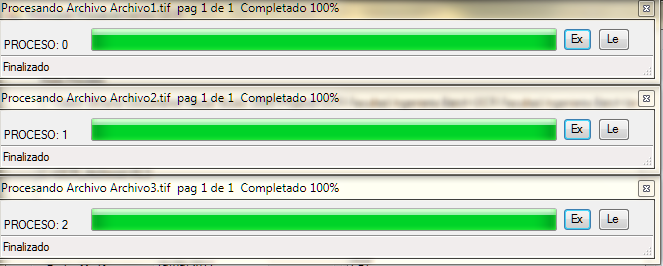


**Autores: Silvana Betancourt, Miguel Ponce**

* **Detener*:*** Permite detener el lote de documentos en proceso.
* **Procesar*:*** Procesa los archivos con los distintas extensiones seleccionadas anteriormente bajo la ruta definida, muestra un cuadro de dialogo que indica el progreso del Ocr y la actualización de registros en la base de datos

Nuestra por cada proceso definido anteriormente una venta de procesamiento como se observa en la figura siguiente

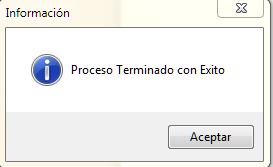
Gráfico 66. Procesamientos de archivos



**Autores: Silvana Betancourt, Miguel Ponce**

Terminado este proceso muestra un mensaje indicando el estado de finalización del Ocr como se muestra a continuación.

Gráfico 67. Terminación del proceso

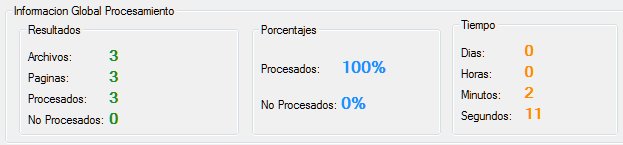


**Autores: Silvana Betancourt, Miguel Ponce**

Por último muestra un resumen de resultados finales como:

* Resumen de resultados, porcentajes y tiempos de respuesta

Gráfico 68. Resumen resultado por lote de documentos en proceso



**Autores: Silvana Betancourt, Miguel Ponce**

**Resultados:** Resumen de la cantidad de documentos procesados o no procesados.

* **Archivos:** Número de archivos para procesamiento.
* **Páginas:** Número de páginas procesadas.
* **Procesados:** Número de documentos procesados, tomados como éxito al cumplir con las condiciones óptimas para el procesamiento.
* **No Procesados:** Número de documentos no procesados, tomados como fracaso debido a inconsistencias en sus parámetros de digitalización, su valor puede verse afectado por el proceso de reprocesamiento.

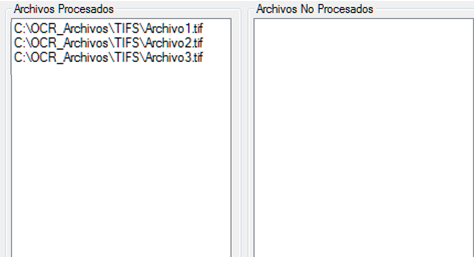
**Porcentaje:** Resumen del porcentaje de documentos para procesamiento o procesados, se toma como fracaso incluso todos los documentos que se encuentran en la carpeta no procesados más lo que se encuentran en un estado de espera.

* + **Procesados:** Porcentaje documentos procesados.
  + **No procesado:** Porcentaje documentos no procesados.

**Tiempo:** Cantidad de días horas, minutos y segundos destinados al procesamiento de un lote de documentos mientras este no sea cancelado, caso contrario su valor inicial será tomado a partir de la hora de procesamiento actual.

* Resumen de archivos procesados: listado con su correspondiente path de los archivos procesados y no procesados.

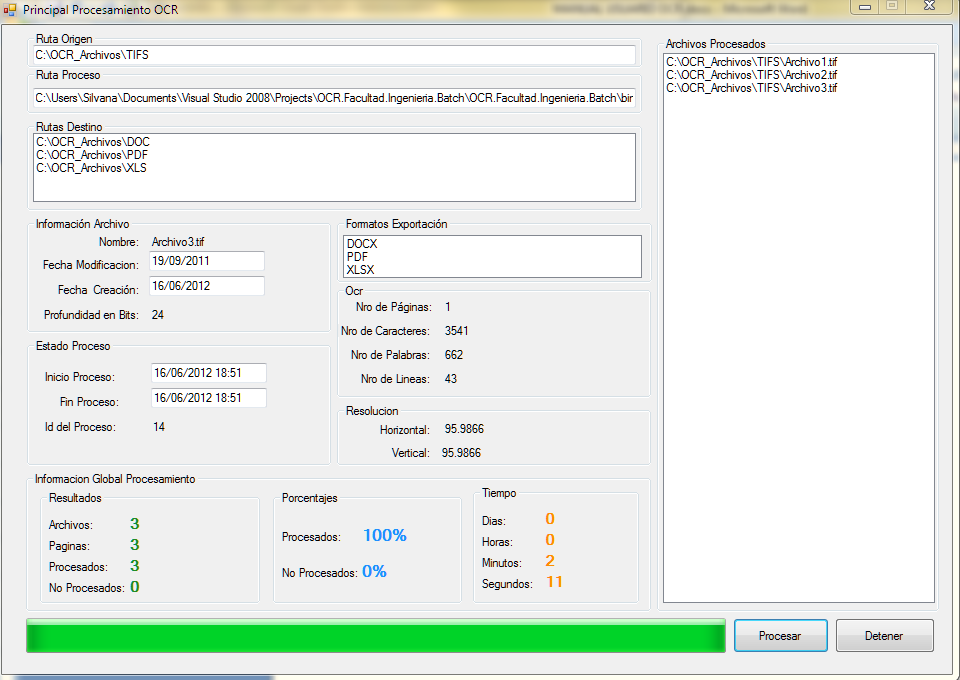
Gráfico 69. Archivos procesados



**Autores: Silvana Betancourt, Miguel Ponce**

* Rutas Destino: muestra un listado de todas las rutas destinos que se definió para los archivos que se realizó el Ocr con los distintos formatos existentes.

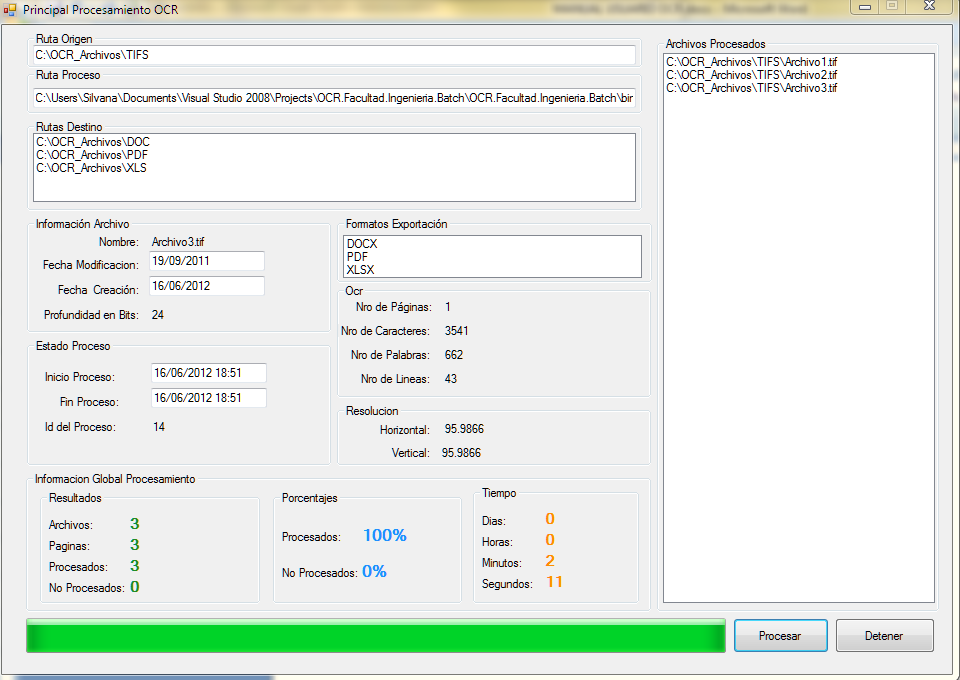
Gráfico 70. Rutas destino



**Autores: Silvana Betancourt, Miguel Ponce**

* Información de Archivo: muestra información referente al nombre del archivo, fecha de modificación, fecha de creación y profundidad en bits que son características propias del archivo

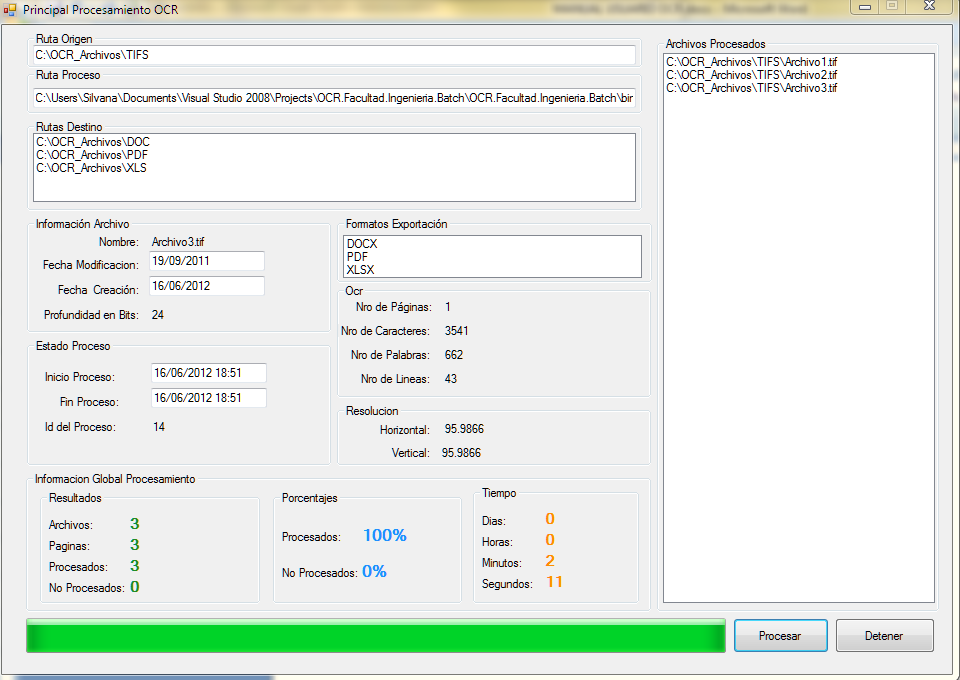
Gráfico 71. Información archivo.

****

**Autores: Silvana Betancourt, Miguel Ponce**

* Estado Proceso: muestra el inicio del proceso, fin del proceso y el identificador del proceso

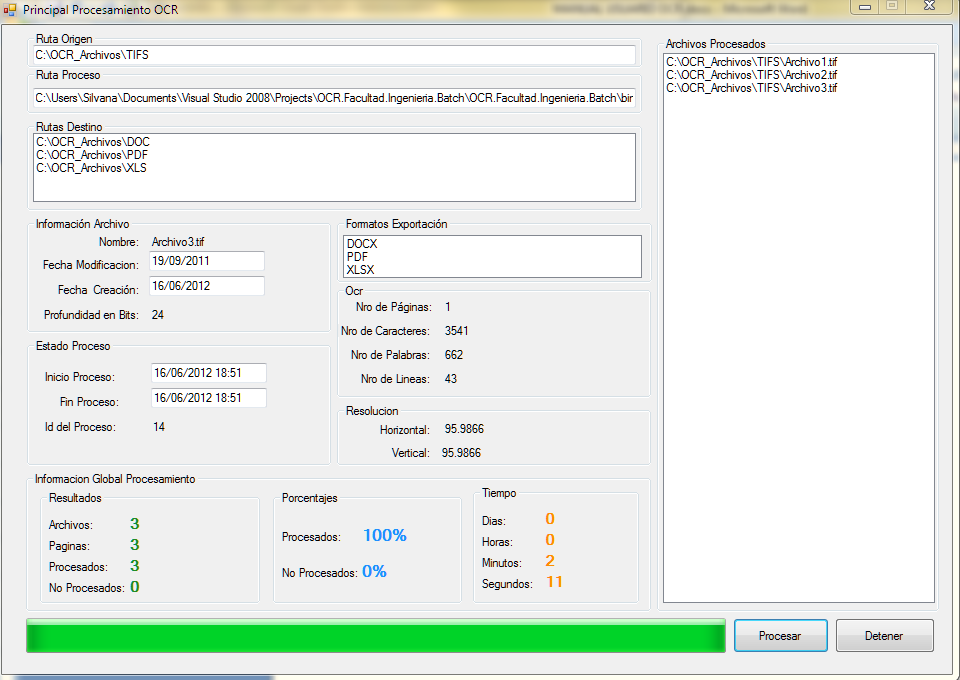
Gráfico 72. Estado proceso

****

**Autores: Silvana Betancourt, Miguel Ponce**

* Formatos de Exportación**:** lista los formatos a los que se realizó el Ocr

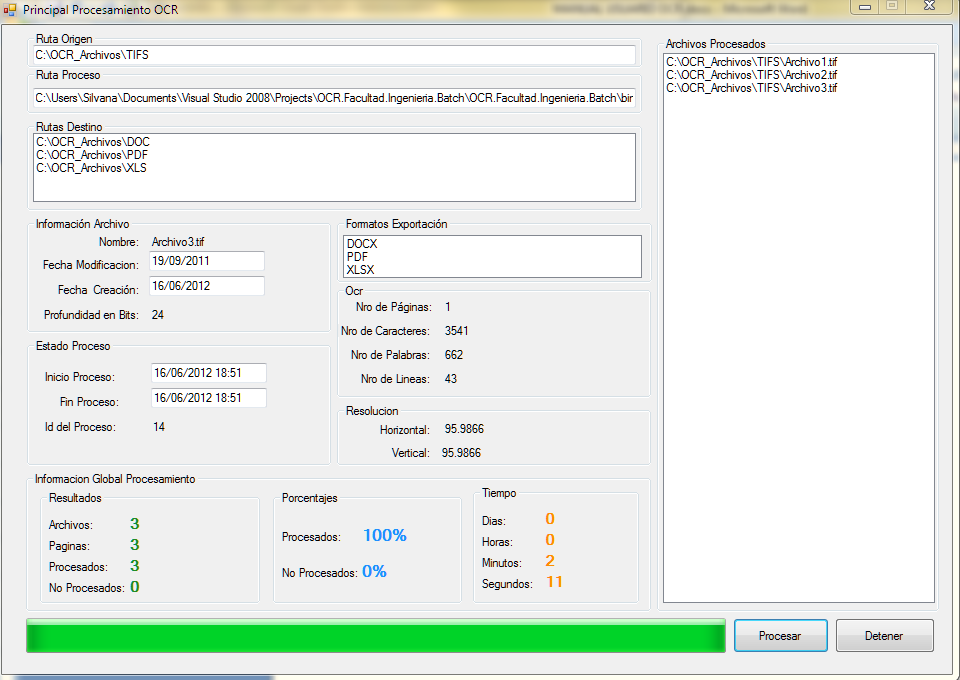
Gráfico 73. Formatos exportación

****

**Autores: Silvana Betancourt, Miguel Ponce**

* Ocr:muestra información respecto a número de páginas, numero de caracteres, numero de palabras y número de líneas que realizo el Ocr

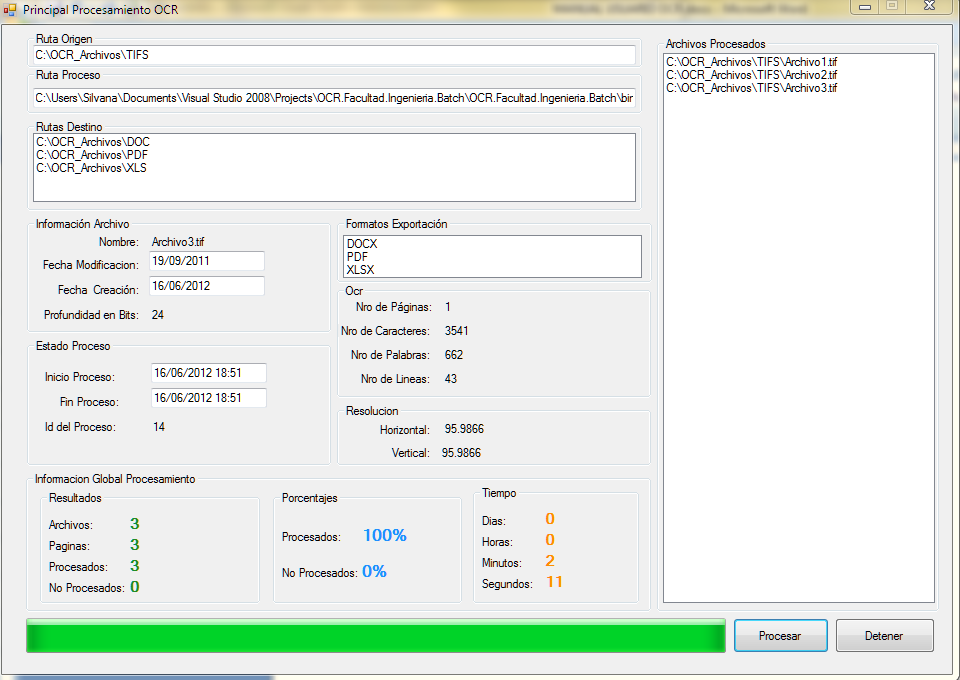
Gráfico 74. Información Ocr

****

**Autores: Silvana Betancourt, Miguel Ponce**

* Resolución:muestra la resolución que tiene la imagen en bits

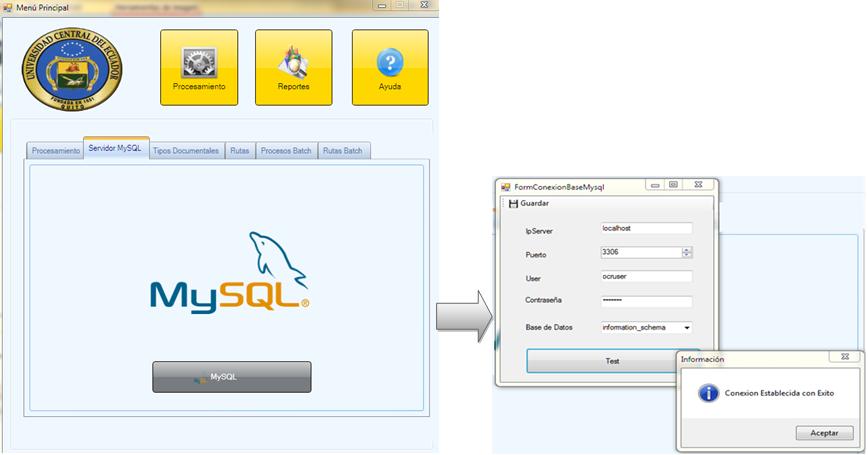
Gráfico 75. Resolución imagen

****

**Autores: Silvana Betancourt, Miguel Ponce**

**Tab Servidor**: Permite conectarse con una base MySql que este instalada en cualquier servidor, para ello debe ingresar la información correspondiente al servidor como se visualiza a continuación.

Gráfico 76. Pantalla tab servidor



**Autores: Silvana Betancourt, Miguel Ponce**

* IpServer: caja de texto que permite ingresar la dirección IP del servidor de base de datos, por defecto es localhost.
* Puerto: es un número, que indica el puerto mediante el cual se conectara a la base de datos, además posee una barra que permite el incremento y decremento en una unidad el valor del campo.
* Base: nombre de la base de datos, este campo acepta todo tipo de caracteres incluidos caracteres especiales.
* User: nombre de usuario registrado y autorizado en la base de datos MySql.
* Contraseña: contraseña del usuario.

Botón test en la parte inferior de la ventana conexión MySql

Gráfico 77. Botón ventana conexión MySql

#### 

**Autores: Silvana Betancourt, Miguel Ponce**

Test:permite testear la conexión con la base de datos, mostrando mensaje sobre el éxito o fracaso de la misma.

Botón Guardar en la parte superior de la ventana conexión MySql

Gráfico 78. Botón guardar

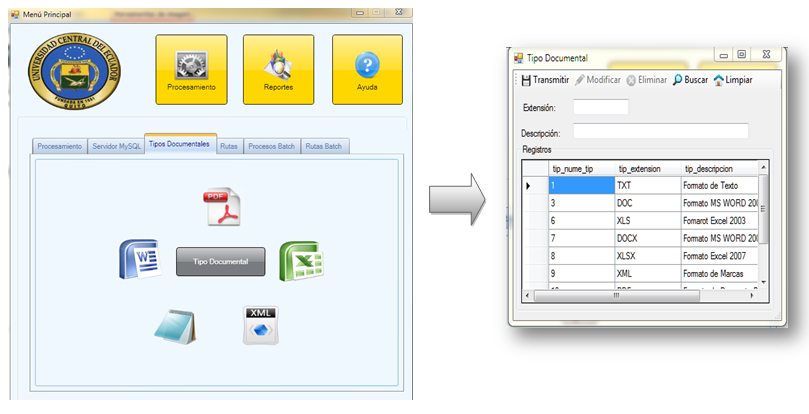
****

**Autores: Silvana Betancourt, Miguel Ponce**

Guardar: Este botón permite guardar la configuración ingresada en los campos de la ventana conexión MySql, previa verificación del estado de conexión con la base.

**Tab Tipos Documentales:** En este tab visualizara una lista de las extensiones y una descripción de formatos existentes, permitiendo guardar, modificar, eliminar, buscar un tipo documental como se observa.

Gráfico 79. Tipo documental



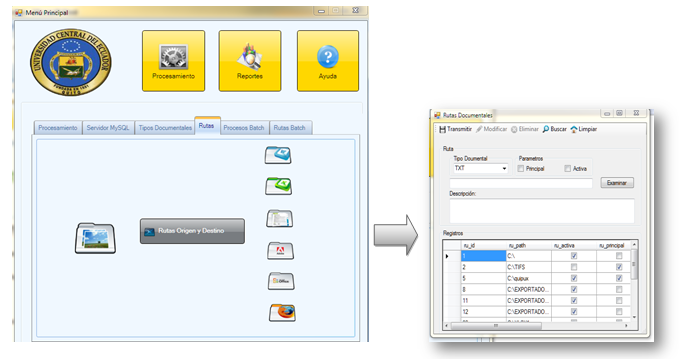
**Autores: Silvana Betancourt, Miguel Ponce**

**Tab Rutas:** Creación de rutas origen y destino, aquí selecciona el tipo documental, parámetros donde especifica si la ruta a definir es principal y activa.

Para definir una ruta se debe tener en cuenta los siguientes puntos:

* Ruta : dirección donde van a ser exportados los archivos que se ha realizado el Ocr
* Ruta principal: dirección donde están las imágenes tif (rutas origen), de las cuales se va a realizar el Ocr
* Ruta Activa: dirección o direcciones (dirección destino) a las cuales se va exportar los archivos que se ha realizado el Ocr, pueden existir varias rutas activas
* Si no existe una ruta activa no se exportaran los archivos que son el resultado de realizar el Ocr, entonces debe haber por lo menos una ruta activa.

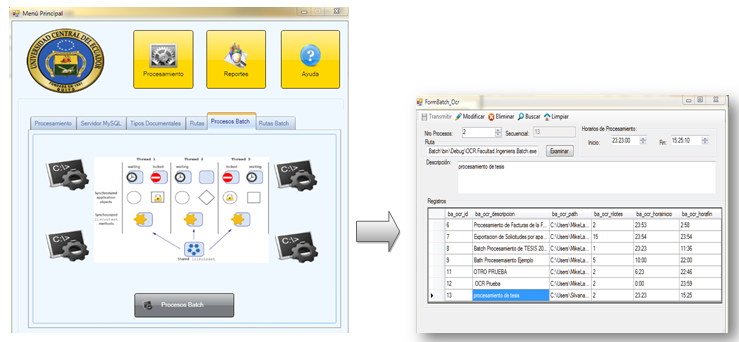
Gráfico 80. Rutas



**Autores: Silvana Betancourt, Miguel Ponce**

**Tab Procesos Batch:** En esta sección se define cuantos procesos se van a ejecutar con el proceso batch que se carga en la sección ruta, el **OCR.Facultad.Ingenieria.Batch.exe** es el encargado de realizar el Ocr.

Gráfico 81. Proceso batch



**Autores: Silvana Betancourt, Miguel Ponce**

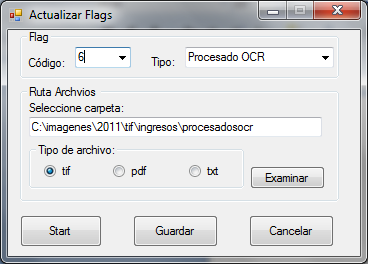
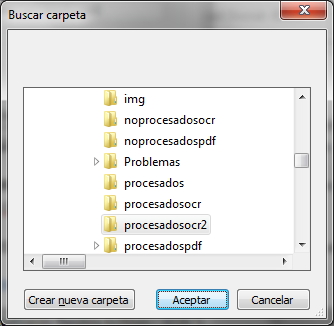
El botón, abre un cuadro dialogo el cual permite la selección del proceso batch.

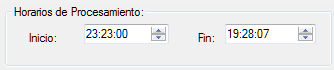
Ilustración 117. Ventana buscar proceso batch



**Autores: Silvana Betancourt, Miguel Ponce**

En las horas de procesamiento se debe definir un intervalo de tiempo en el que se puede realizar el procesamiento para ello se debe ingresar la hora de inicio y fin como se observar.

Gráfico 82. Apartado utilizado para el procesamiento por horarios

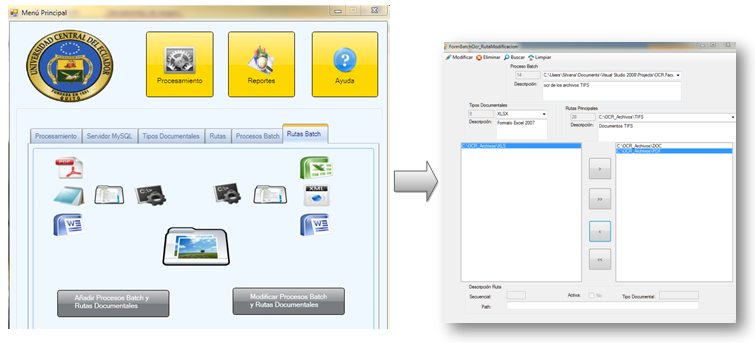


**Autores: Silvana Betancourt, Miguel Ponce**

**Rutas Batch:** Selección de tipo documental, proceso batch y rutas principales definidos anteriormente.

En la sección de tipo documental se debe añadir las rutas secundarias existentes en cada tipo documental hacia el lado derecho de la ventana como se visualiza en la figura siguiente.

Gráfico 83. Rutas batch



**Autores: Silvana Betancourt, Miguel Ponce**

**Botón Reportes:** Visualizara un conjunto de tabs que se describe más adelante

Gráfico 84. Botón reportes



**Fuente: Silvana Betancourt, Miguel Ponce**

Tabs de Reportes

Gráfico 85. Tabs de la pantalla principal reportes



**Autores: Silvana Betancourt, Miguel Ponce**

**Tab Rutas:** Aquí se visualizara un conjunto de reportes como: rutas principales, rutas activas, rutas secundarias y rutas activas.

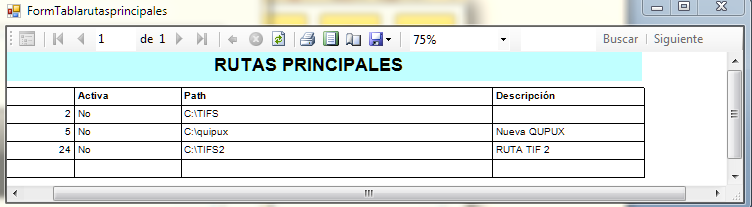
Gráfico 86. Reporte rutas

****

**Autores: Silvana Betancourt, Miguel Ponce**

* **Rutas Principales.-** Listado de rutas principales que han sido creadas.

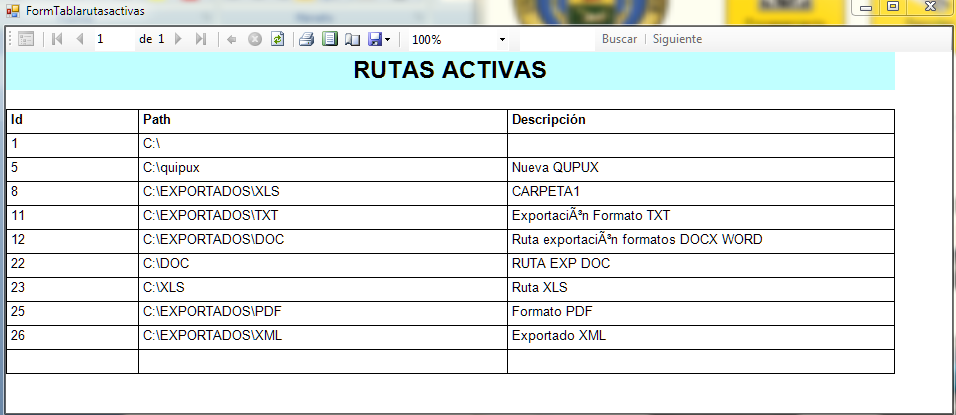
Gráfico 87. Rutas principales

****

**Autores: Silvana Betancourt, Miguel Ponce**

* **Rutas Activas.-** Listado de rutas activas que se ha definido en el sistema

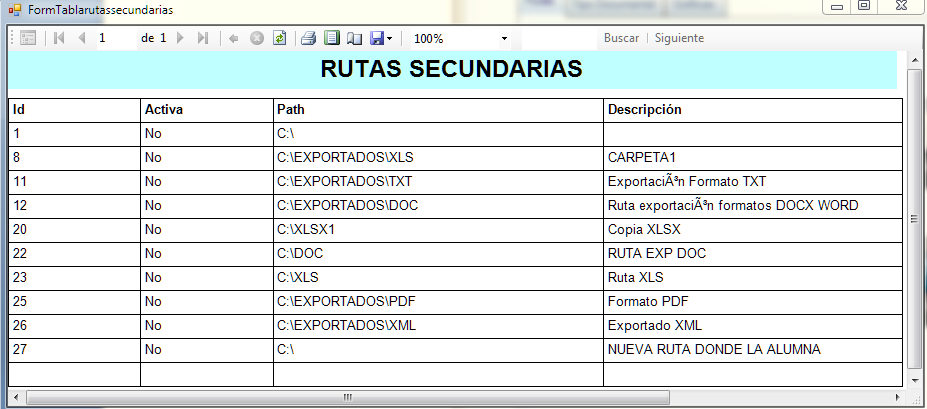
Gráfico 88. Rutas activas

****

**Autores: Silvana Betancourt, Miguel Ponce**

* **Rutas Secundarias.-** Listado de rutas secundarias que se han creado en el sistema.

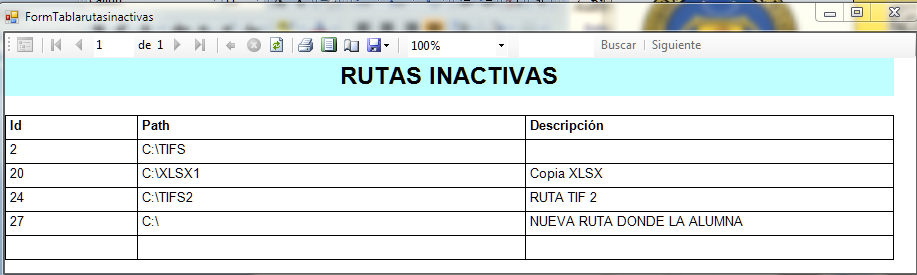
Gráfico 89. Rutas secundarias

****

**Autores: Silvana Betancourt, Miguel Ponce**

* **Rutas Inactivas.-** Listado de rutas inactivas existentes en el sistema

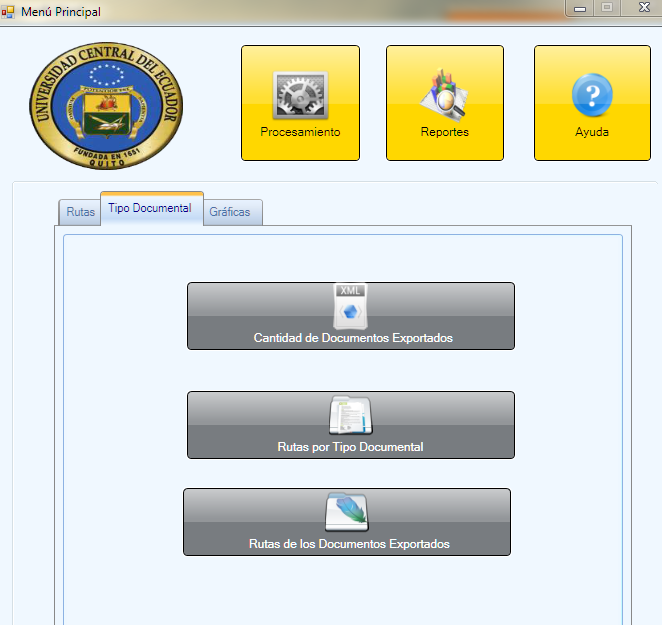
Gráfico 90. Listado rutas inactivas



**Autores: Silvana Betancourt, Miguel Ponce**

**Tab Tipo documental:** Aquí se visualizara un conjunto de reportes como: cantidad de documentos exportados, rutas por tipo documental y rutas de los documentos exportados.

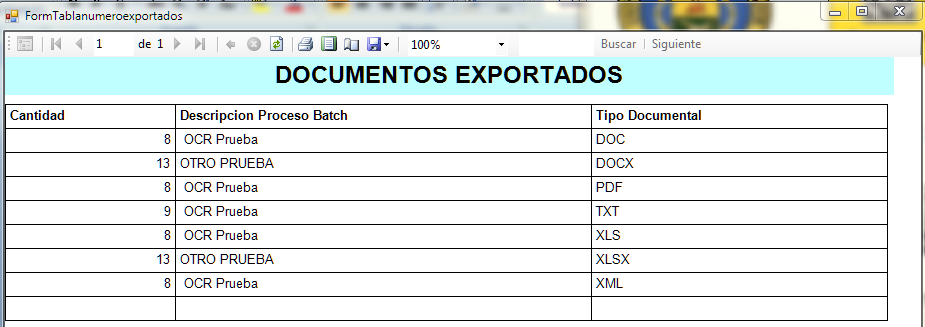
Gráfico 91. Tipo documental



**Autores: Silvana Betancourt, Miguel Ponce**

* Cantidad de documentos exportados: Listado por tipos documentales que han sido procesados y exportados.

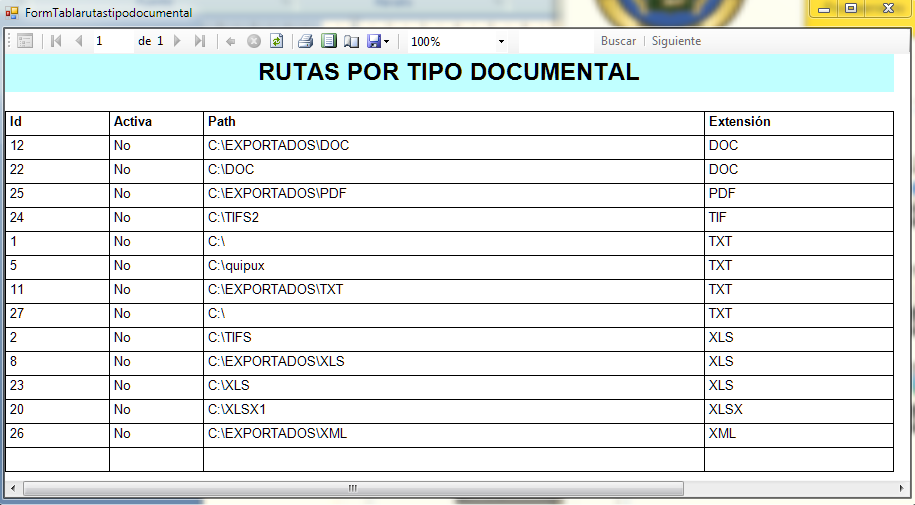
Gráfico . Documentos exportados



**Autores: Silvana Betancourt, Miguel Ponce**

* Rutas por tipo documental: Listado de rutas creadas por tipo documental definido

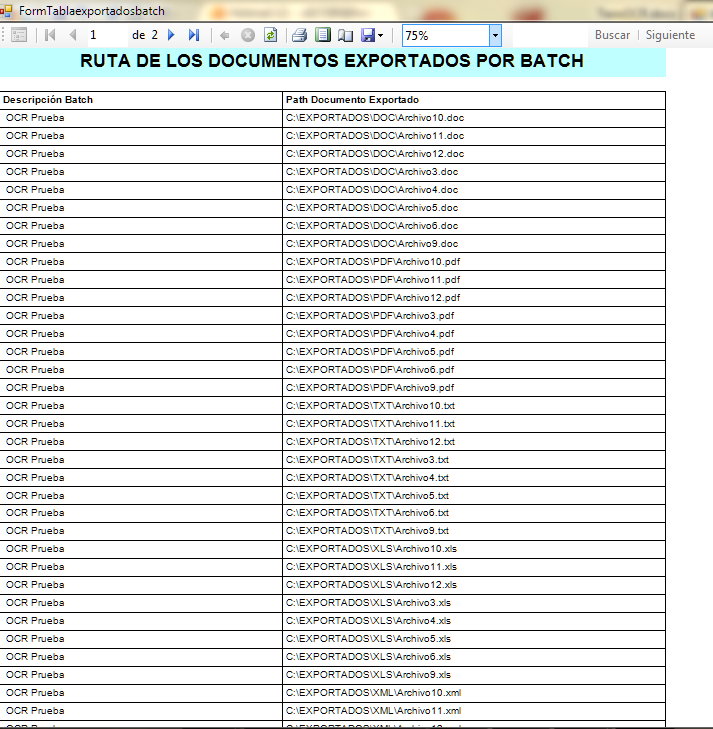
Gráfico 93. Rutas por tipo documental



**Autores: Silvana Betancourt, Miguel Ponce**

* Rutas de los documentos exportados: Listado de batch por rutas de los tipos documentales definidos en el sistema.

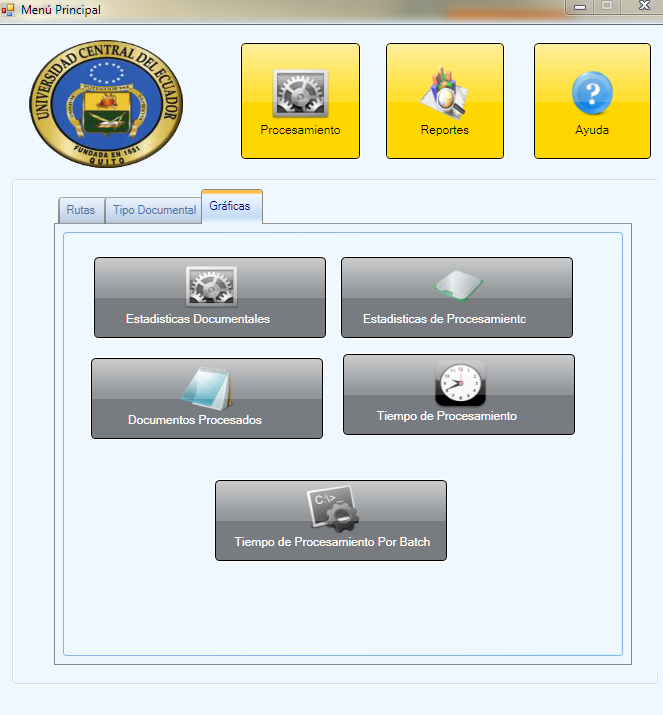
Gráfico 94. Listado de procesos batch



**Autores: Silvana Betancourt, Miguel Ponce**

**Tab Graficas:** Aquí visualizara un conjunto de reportes como: estadísticas documentales, estadísticas de procesamiento, documentos procesados y tiempo de procesamiento.

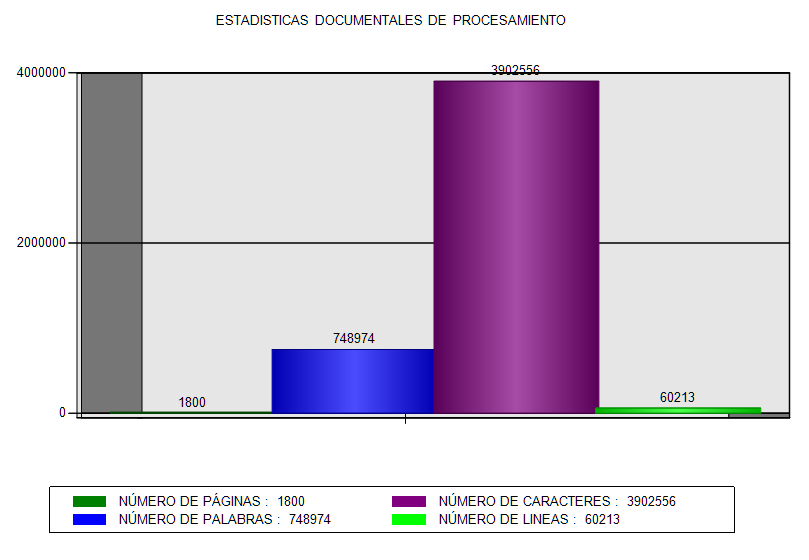
Gráfico 95. Estadísticas de procesamiento



**Autores: Silvana Betancourt, Miguel Ponce**

* Estadísticas Documentales:Nuestra estadísticas de número de páginas procesadas, número de palabras procesadas, numero de caracteres procesados y número de líneas que se procesaron cuando se realizó el Ocr

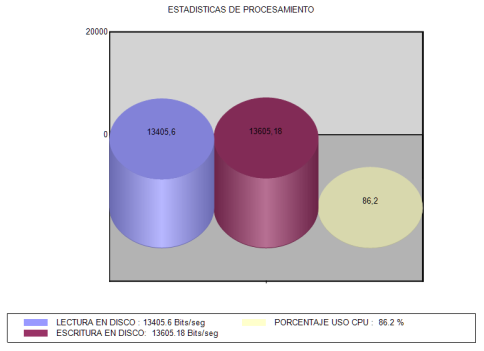
Gráfico 96. Estadísticas documentales

****

**Autores: Silvana Betancourt, Miguel Ponce**

* Estadísticas de Procesamiento: Este reporte muestra características como tiempo de lectura en disco, tiempo de escritura en disco, porcentaje de uso CPU, estos datos ayudan a tener una idea del tiempo y lo que está consumiendo el sistema al realizar el Ocr, permitiendo de esta manera tomar medidas para prevenir cualquier anomalía.

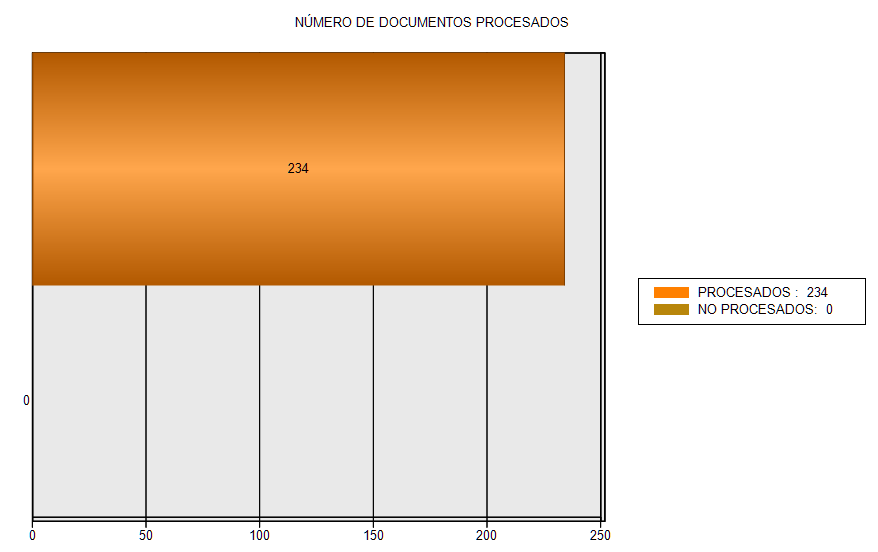
Gráfico 97. Estadística de procesamiento



**Autores: Silvana Betancourt, Miguel Ponce**

* Documentos procesados: Este reporte permite conocer la cantidad de documentos procesados y no procesados.

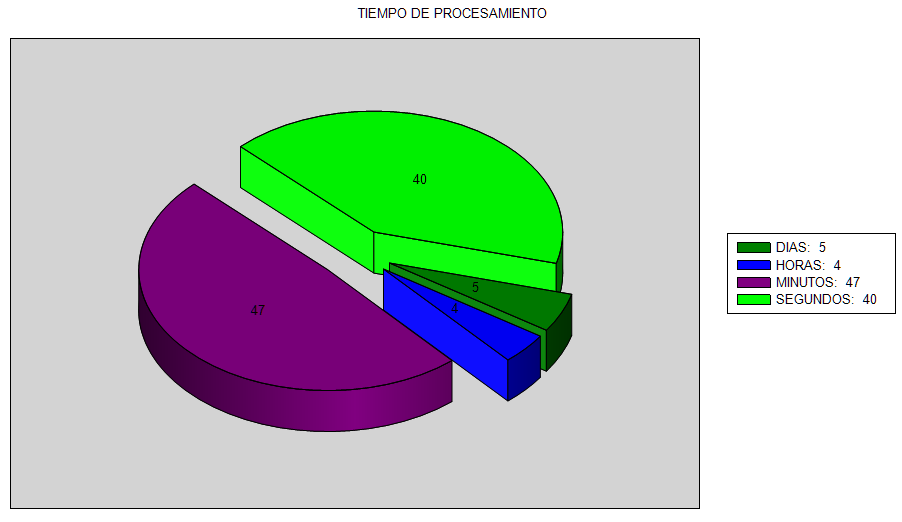
Gráfico 98. Número de documentos procesados

****

**Autores: Silvana Betancourt, Miguel Ponce**

* Tiempo de procesamiento: Este reporte permite conocer el tiempo de procesamiento en días, horas, minutos y segundos.

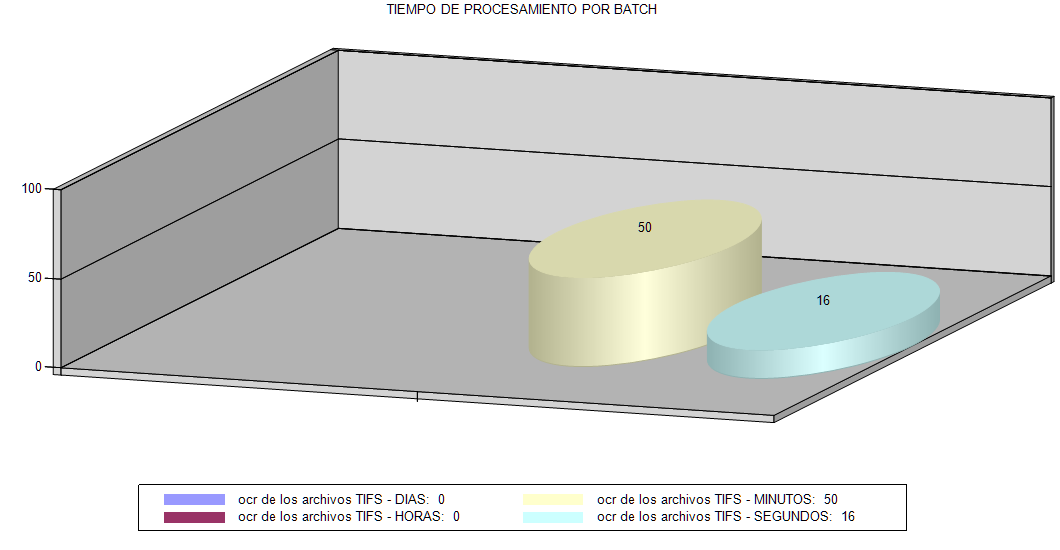
Gráfico 99. Tiempo de procesamiento



**Autores: Silvana Betancourt, Miguel Ponce**

* Tiempo de procesamiento por batch: Este reporte permite al usuario conocer cuántos batch se procesaron y cuánto tiempo se demoraron en días, horas, minutos y segundos, en procesar los documentos de formato tif.

Gráfico 100. Tiempo de procesamiento por batch



**Autores: Silvana Betancourt, Miguel Ponce**

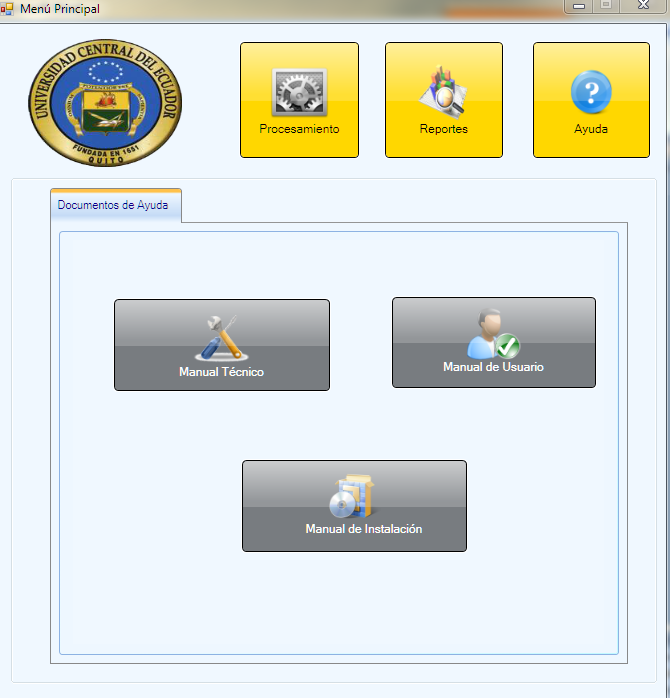
**Botón Ayuda:** Aquí se visualizara tres documentos de ayuda como son: manual de usuarios, manual técnico y manual de instalación.

Gráfico 101. Ayuda



**Autores: Silvana Betancourt, Miguel Ponce**

Gráfico 102. Documentos de ayuda

****

**Autores: Silvana Betancourt, Miguel Ponce**