DetectaNIF: Manual de uso

# Acerca de

*DetectaNIF* es una utilidad para localizar ficheros candidatos a contener datos de carácter personal. Para ello, busca la presencia de cadenas de caracteres coincidentes con el formato de NIF/NIE de persona física.

# Modo de funcionamiento

*DetectaNIF* busca recursivamente en todas las carpetas que se le hayan indicado. Para cada fichero existente en cada carpeta determina:

* Si tiene alguna de las extensiones de fichero configuradas para ser elegible.
* Si tiene el tamaño mínimo configurado para ser elegible.
* Si el tiempo transcurrido desde su creación supera el mínimo configurado para ser elegible.

En caso afirmativo, el fichero es incluido en una cola para su exploración. La cola de ficheros es atendida por un número de hilos concurrentes (configurable). Cada hilo toma un fichero, explora su contenido y escribe el resultado en el fichero de salida con formato CSV (codificado en UTF-8). La exploración tiene estas características:

* Es binaria (no necesita aplicaciones externas para abrir cada fichero).
* Busca cadenas coincidentes con el formato de NIF/NIE con codificación ANSI, UTF-8 o UTF-16.
* Si es posible, se preserva la fecha de último acceso de cada fichero.
* No se exploran los últimos 17 bytes del fichero, puesto que no pueden contener una cadena de NIF/NIE codificada en UTF-16, pero podría contener una cadena NIF/NIE en UTF-8/ANSI. Por tanto, hay un margen de error de un NIF.
* En ficheros comprimidos es posible detectar las cadenas NIF/NIE, pero su número podría ser inferior al real.
* No arroja resultados en ficheros cifrados, si bien, no está afectado por el cifrado de volumen (tipo “BitLocker”).

La exploración tiene dos límites (configurables) con el propósito de disminuir su tiempo de ejecución. En primer lugar, se finaliza la exploración cuando ya se ha alcanzado un cierto número de cadenas NIF/NIE (parámetro “MaxNIF”). Por otra parte, se finaliza la exploración si la primera porción del fichero no contiene ninguna cadena NIF/NIE:

* La porción en concreto se calcula como un porcentaje del tamaño del fichero, configurado en el parámetro “MaxSinNIF”.
* No obstante, siempre se procesan bloques enteros de 64 kilobytes.

## Rutas de fichero y carpeta

*DetectaNIF* es capaz de manejar rutas largas de fichero (hasta 65535 bytes) y en formato UNICODE. Son válidas las rutas UNC (“\\servidor\recurso”). Se recomienda usar rutas absolutas.

# Parámetros de exploración (configuración)

*DetectaNIF* admite el suministro de los parámetros de exploración por tres vías:

* Argumentos de la línea de comandos (admite sólo los parámetros “Salida” y “Carpeta”)
* Fichero de configuración indicado por línea de comandos (admite todos los parámetros)
* Fichero de texto plano indicado por línea de comandos (sólo para carpetas a explorar, parámetro “Carpeta”).

Todos los parámetros se identifican por una etiqueta. Sea cual sea la vía de suministro, los parámetros mínimos necesarios para una exploración son:

* Una carpeta a explorar (parámetro “Carpeta”)
* Un fichero de salida (parámetro “Salida”).

Los demás parámetros son opcionales. En caso de omisión, se utilizará un valor por defecto.

## Descripción de los parámetros

### Parámetro “GenerarResumen”

Indica si se desea generar un fichero adicional a modo de resumen de la exploración (recomendado). El fichero tendrá el mismo nombre y ubicación que el fichero de salida (parámetro “Salida”) pero con extensión “.log”.

Valores admitidos: “true” o “false” (valor por omisión).

No es obligatorio. Suministrado por fichero de configuración.

### Parámetro “IncluirErrores”

Indica si se desea dejar constancia, en el fichero de salida, de los ficheros que no pudieron ser explorados a causa de cualquier error de acceso al fichero. Generalmente, estos errores se deben a insuficiencia de permisos.

Valores admitidos: “true” (valor por omisión) o “false”.

No es obligatorio. Suministrado por fichero de configuración.

### Parámetro “NumeroHilos”

Indica el número de exploraciones concurrentes.

Valores admitidos: número entero entre uno y diez, ambos inclusive. Por omisión, seis hilos.

Se recomienda hacer coincidir este parámetro con el número de núcleos físicos del microprocesador. No es obligatorio. Suministrado por fichero de configuración.

### Parámetro “MinDiasCreado”

Indica el número mínimo de días que han de transcurrir desde la fecha de creación de un fichero para que éste sea elegible para su exploración.

Valores admitidos: entero mayor o igual que cero (valor por omisión). Por omisión, 365 días.

No es obligatorio. Suministrado por fichero de configuración.

### Parámetro “MinBytesTamaño”

Indica el tamaño mínimo que ha de tener un fichero para que sea elegible para su exploración.

Valores admitidos: entero mayor o igual que cero. Por omisión, 3 megabytes.

No es obligatorio. Suministrado por fichero de configuración.

### Parámetro “MaxNIF”

Indica el número máximo de cadenas coincidentes con el formato de NIF/NIE que serán contabilizadas. Superado este límite, se finalizará la exploración del fichero.

Valores admitidos: entero mayor o igual que veinte. Por omisión, 1000.

No es obligatorio. Suministrado por fichero de configuración.

### Parámetro “MaxSinNIF”

Indica, en porcentaje, la porción inicial del fichero que debe contener alguna cadena NIF/NIE para continuar con la exploración.

Valores admitidos: entero entre 40 y 100, ambos inclusive. Estos límites evitan falsos negativos. Por omisión, cincuenta por ciento.

No es obligatorio. Suministrado por fichero de configuración.

### Parámetro “Extensiones”

Indica las extensiones que puede tener un fichero para que sea elegido para su exploración.

Suministrado por fichero de configuración. No es obligatorio. Por omisión se emplean las extensiones más comunes de bases de datos ofimáticas:

Extensiones=xls accd? md? csv txt xlsx\* db\* prn dif tab adp gdb ib jet odb sbf sqlite\* xml

Todas las extensiones deben indicarse en una sola línea, separadas por un espacio y sin punto. Se admiten los caracteres comodines (asterisco e interrogación). Por ejemplo, para explorar todos los ficheros, puede indicarse:

Extensiones=\*

### Parámetro “Salida”

Indica el nombre y ruta del fichero a generar con los resultados. El fichero tendrá formato CSV codificado en UTF-8. Obligatorio, si bien, suministrado tanto por fichero de configuración como línea de comandos. Si se suministran ambos, solamente se tendrá en cuenta la línea de comandos.

## Parámetro “Carpeta”

Indica la ruta a una carpeta a explorar (recursivamente). Suministrado por fichero de configuración, línea de comandos o fichero de texto plano. Se tendrán en cuenta todos los valores suministrados y por cualquier vía. Por tanto, es posible configurar varias carpetas. Alguna de las carpetas configuradas puede ser ignorada si:

* No es una ruta bien formada a una carpeta
* Se ha configurado otra carpeta que la contiene
* La carpeta no existe.

Ejemplo mediante fichero de configuración:

Carpeta=e:\recursos

Carpeta=l:\recursos

Ejemplo mediante fichero de texto plano:

e:\recursos

l:\recursos

Ejemplo mediante línea de comandos:

detectaNIF –carpeta “e:\recursos” “l:\recursos”

# Formato del fichero de configuración

Todos los parámetros de exploración se pueden suministrar en un fichero de configuración con formato de texto plano. Se recomienda utilizar codificación UTF-8 para evitar que ciertos caracteres sean irreconocibles (por ejemplo, vocales acentuadas). Puede generarlo con la utilidad “notepad”, opción “guardar como”. En el diálogo de selección de fichero elija la codificación UTF-8.

Incluya un parámetro en cada línea de texto con el siguiente formato:

Nombre=valor

No inserte espacios en blanco innecesarios y finalice con un salto de línea. Las líneas que no contienen el carácter de igualdad (=) son ignoradas, por tanto, se pueden introducir líneas a modo de comentario. Por ejemplo:

# FICHERO DE CONFIGURACIÓN DE EJEMPLO

GenerarResumen=true

IncluirErrores=true

NumeroHilos=4

# Un Año

MinDiasCreado=365

MinBytesTamaño=3145728

MaxNIF=1000

MaxSinNIF=50

Extensiones=xls xlsx

Carpeta=L:\recursos

Carpeta=E:\recursos

Salida=C:\temp\candidatos NIF.csv

Puede generar un fichero de configuración por línea de comandos, ejecutando:

detectaNIF –plantilla cfg.txt

A continuación, debe editar el fichero (“cfg.txt” en el ejemplo).

# Línea de comandos

Todos los argumentos de línea de comandos se indican con el siguiente formato:

-Etiqueta valor

De manera opcional y en cualquier orden. Se recomienda encerrar el valor entre comillas.

Las etiquetas admitidas son:

* *Cfg*: indica un fichero de configuración a cargar
* *Plantilla*: indica una plantilla de fichero de configuración a generar
* *Carpeta*: análoga al parámetro de configuración del mismo nombre
* *Salida*: análoga al parámetro de configuración del mismo nombre
* *FicheroCarpetas*: indica el fichero de texto plano que contiene las carpetas a explorar.

Para ejecutar con los parámetros por defecto:

detectaNIF –Salida <fichero> -Carpeta <carpeta> [<carpeta> …]

Por ejemplo:

detectaNIF –Salida “c:\temp\resultado.csv” -Carpeta “\\miservidor.midominio\recursos”

Para tomar los parámetros de un fichero de configuración:

detectaNIF –cfg <fichero>

Para tomar las carpetas a explorar de un fichero de texto plano:

detectaNIF –Salida <fichero resultado> -FicheroCarpetas <fichero texto>

Por ejemplo:

detectaNIF –Salida “resultado.csv” -FicheroCarpetas “C:\temp\carpetasExplorar.txt”

Para obtener una descripción de los argumentos, ejecute sin más:

detectaNIF

# Formato del fichero (opcional) de carpetas a explorar

Se trata de un fichero de texto plano conteniendo las rutas a las carpetas para explorar. Una ruta por cada línea. Se recomienda generar el fichero con codificación UTF-8.

# Requisitos

*DetectaNIF* requiere un sistema operativo Microsoft Windows de 64 bits.

Debe ejecutarse en el contexto de un usuario con privilegios de acceso (en lectura) a las carpetas y ficheros que se pretenden explorar. Para preservar las fechas de último acceso también es necesario el permiso de lectura y escritura de atributos.