



Perfilamiento Digital: Análisis Forense de Conducta mediante SQL

El perfilamiento digital es una metodología sistemática que permite reconstruir el comportamiento, intenciones y nivel técnico de un usuario mediante el análisis estadístico de bases de datos de navegadores. No se trata de especulación, sino de evidencia cuantificable extraída de artefactos digitales concretos.

Fundamentos del Perfilamiento Técnico



Análisis de Frecuencia

Identificación de patrones rutinarios mediante conteo de visitas en bases de datos. Distingue comportamiento habitual de accesos esporádicos.



Intención de Consultas

Extracción y análisis de parámetros de búsqueda en URLs para reconstruir procesos cognitivos y objetivos del usuario.



Contexto del Entorno

Determinación de sofisticación técnica mediante análisis de User-Agent, sistema operativo y herramientas utilizadas.

Este enfoque triangulado proporciona un perfil forense robusto basado en tres pilares complementarios: **qué hace** (frecuencia), **qué busca** (intención) y **cómo lo hace** (entorno técnico). Las fuentes principales son las bases de datos places.sqlite (Firefox) y History (Chrome).

Frecuencia de Visita: Diferenciando el Hábito de lo Casual



Visita Casual

1 hit único

- Acceso puntual sin repetición
- Puede ser accidental o referenciado
- Bajo valor investigativo

Visita Recurrente

>10 hits repetidos

- Patrón habitual establecido
- Alto valor para perfilamiento
- Indica interés sostenido

El campo clave para este análisis es `visit_count`, presente en las tablas `moz_places` (Firefox) o `urls` (Chrome). Este valor cuantifica el número de veces que un usuario ha accedido a una URL específica.

❑ **Indicador Forense:** Los sitios más visitados definen arquetipos conductuales. Sitios de apuestas, trading, pornografía o repositorios técnicos (GitHub, StackOverflow) revelan perfiles psicológicos y técnicos diferenciados.

Intención de Búsqueda: Reconstruyendo el Proceso Mental

Las cadenas de consulta en URLs (Query Parameters) constituyen una ventana directa al proceso cognitivo del usuario. Cada búsqueda capturada es una expresión explícita de necesidad, duda o intención investigativa.

Patrón de Análisis

Identificar URLs que contengan patrones como `google.com/search?q=` o el parámetro `&q=` en cualquier motor de búsqueda. Estos fragmentos contienen las palabras clave exactas introducidas por el usuario.

Búsquedas Técnicas

Ejemplos: "CCenter.exe", "IObit Uninstaller",
"como limpiar registro Windows"

→ Perfil: Usuario técnico con conocimiento
de herramientas de sistema

Búsquedas de Rutas Locales

Ejemplos: "file:///C:/Users/", "AppData Local
Temp"

→ Perfil: Exploración activa del sistema de
archivos

Búsquedas Anti-Forenses

Ejemplos: "borrar logs Windows", "eliminar
historial permanente", "CCleaner"

→ Perfil: Conciencia de rastreabilidad
digital

Entorno Técnico: Decodificando el User-Agent

El User-Agent es una cadena de texto que identifica el navegador, sistema operativo y dispositivo utilizado. Su análisis permite determinar el nivel de sofisticación técnica del usuario y detectar intentos de anonimización.

O1

Identificación del Sistema Operativo

Windows 10, Linux Ubuntu, macOS, Android, iOS

O2

Tipo de Navegador

Chrome, Firefox, Edge, Safari, Tor Browser

O3

Categoría de Dispositivo

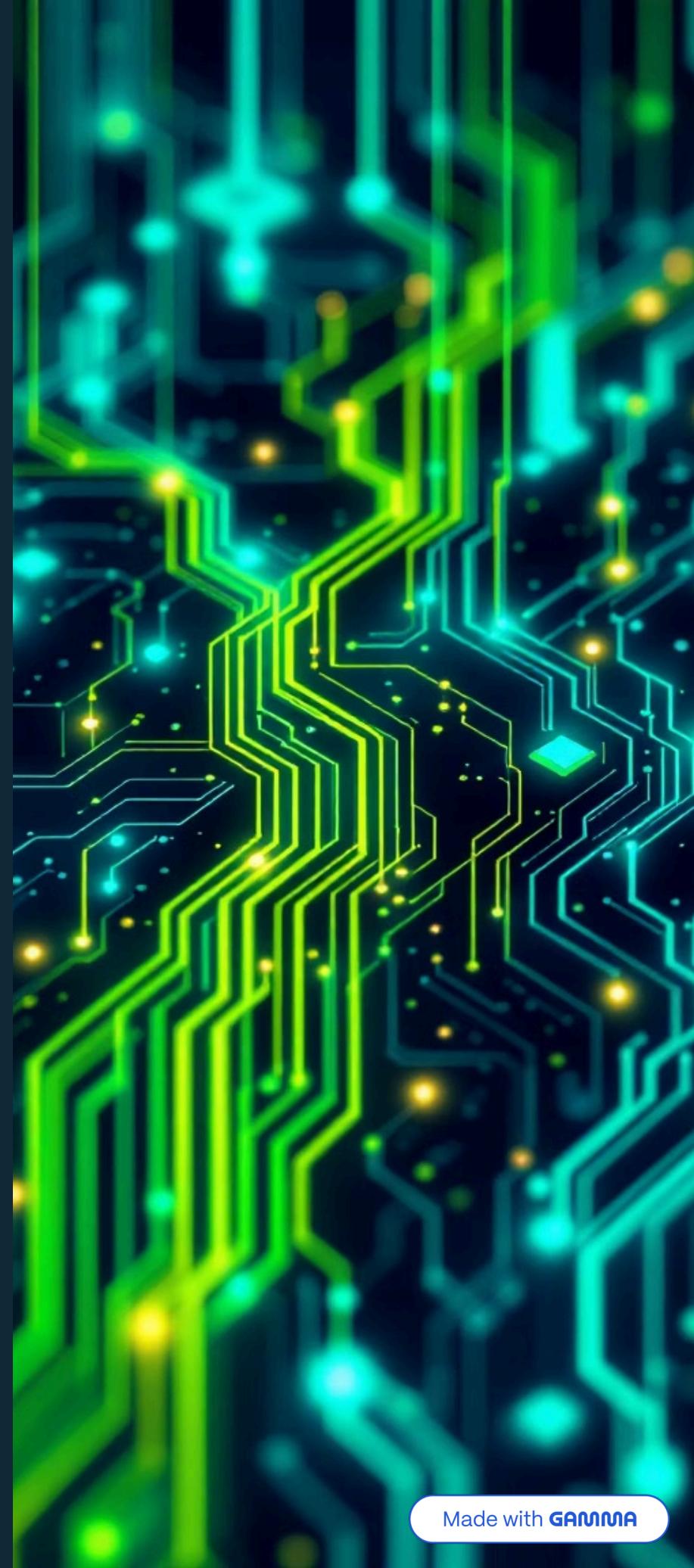
Desktop, móvil, tablet, smart TV

O4

Evaluación de Sofisticación

Uso de Tor, VPN, navegadores especializados

- ▢ **Ejemplo Interpretativo:** Un User-Agent que muestra "Windows NT 10.0; Win64; x64" con "Chrome/120" indica un usuario estándar. La presencia de "Tor Browser" o modificaciones manuales del UA sugieren conocimiento avanzado y posible evasión.



Persistencia de Sesión: Arqueología de Cookies

Artefactos Clave en cookies.sqlite

Las cookies pueden persistir incluso después de que el historial de navegación sea eliminado, convirtiéndose en evidencia residual valiosa.

Indicadores Forenses

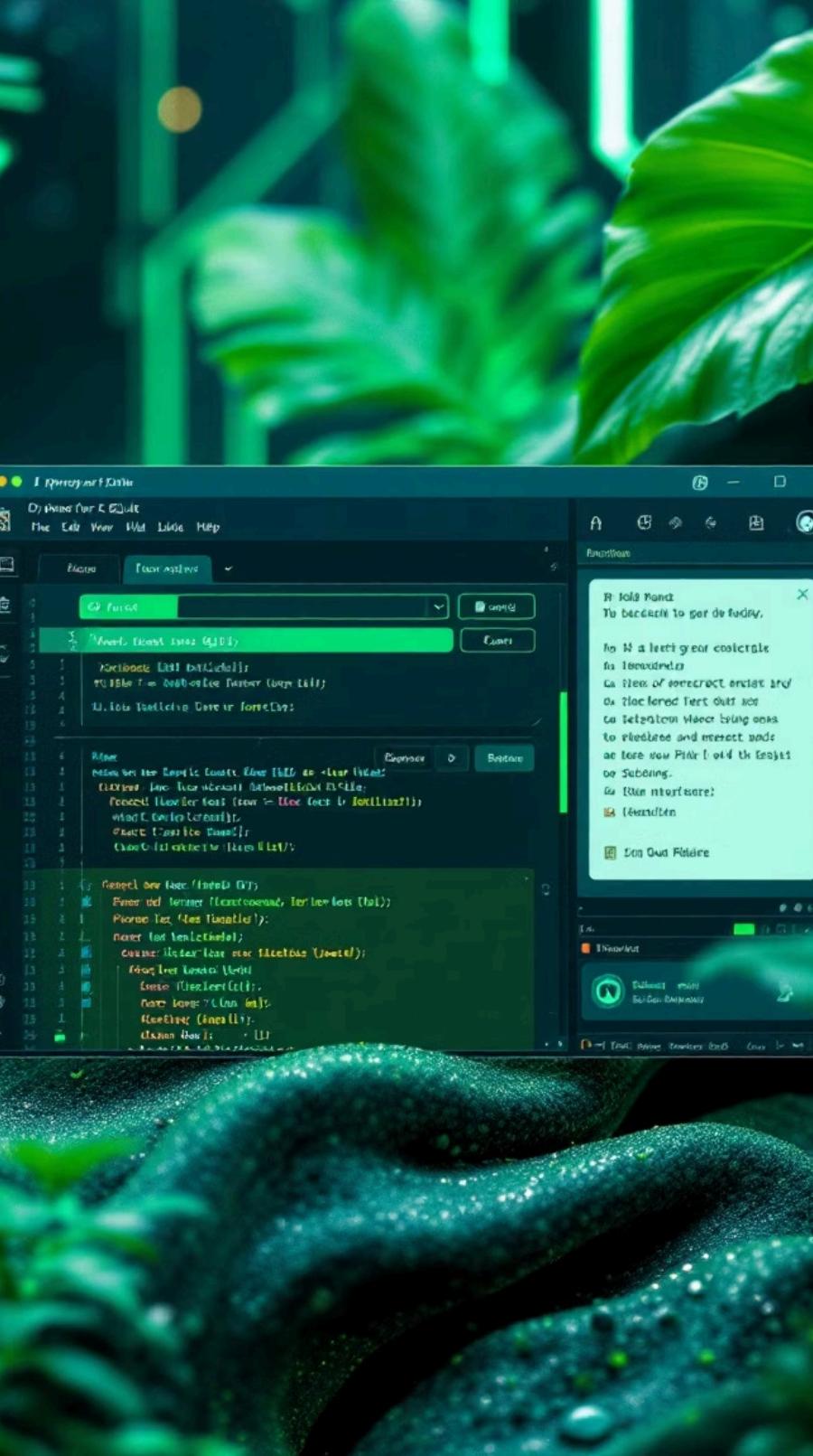
- **Cookies huérfanas:** Dominios presentes en cookies pero ausentes en historial
- **Fechas de expiración lejanas:** Cookies configuradas para persistir años
- **Cookies de sesión persistentes:** Indican sesiones activas no cerradas correctamente

Las cookies de LinkedIn, Clarín u otros servicios pueden revelar actividad no reflejada en el historial visible, especialmente tras intentos de limpieza.



La comparación cruzada entre dominios en cookies.sqlite y URLs en el historial permite identificar brechas temporales, sesiones borradas selectivamente o intentos de ofuscación de la actividad digital.

Configuración del Entorno



Herramienta Requerida

DB Browser for SQLite - Software open source para exploración de bases de datos SQLite sin necesidad de línea de comandos.

Archivos Objetivo

History (Chrome/Edge) ubicado en
%LocalAppData%\Google\Chrome\User Data\Default\
places.sqlite (Firefox) en %AppData%\Mozilla\Firefox\Profiles\

Preparación del Análisis

Crear copia forense: Nunca trabajar sobre el archivo original. El proceso del navegador puede bloquear el acceso (error "Database Locked").

Cierre de Procesos

Asegurar que el navegador esté completamente cerrado, incluyendo procesos en segundo plano visibles en el Administrador de Tareas.

Query de Frecuencia de Visitas

Consulta SQL para Identificación de Hábitos

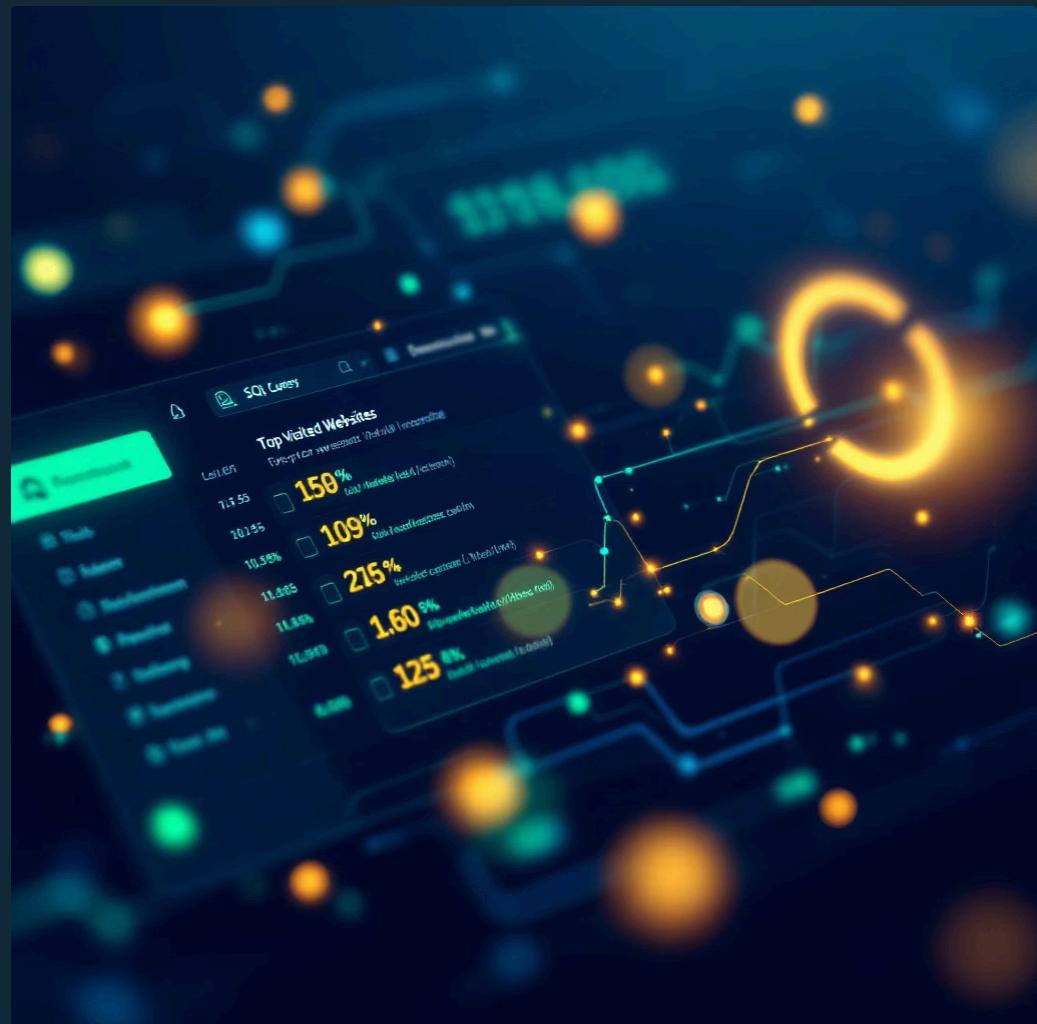
```
SELECT url, title, visit_count  
FROM urls  
WHERE visit_count > 0  
ORDER BY visit_count DESC  
LIMIT 20;
```

Esta query extrae las 20 URLs más visitadas, ordenadas de mayor a menor frecuencia. El campo `visit_count` es el indicador primario de conducta rutinaria.

Interpretación de Resultados

Los **top 5 resultados** constituyen la "Firma de Conducta" del usuario:

- Redes sociales: Perfil social/comunicativo
- Sitios de noticias: Interés informativo
- Foros técnicos: Perfil especializado
- Entretenimiento: Hábitos de ocio
- Banca/Trading: Actividad financiera



- Nota Práctica:** Filtrar resultados con `visit_count > 10` para eliminar ruido de visitas casuales y enfocarse en comportamiento recurrente significativo.

Extracción de Búsquedas

La reconstrucción de búsquedas permite visualizar el proceso mental del usuario, revelando necesidades, dudas, intenciones e incluso premeditación de actividades específicas.

Query SQL para Búsquedas

```
SELECT url, title, last_visit_time  
FROM urls  
WHERE url LIKE '%search?q=%'  
    OR url LIKE '%&q=%'  
ORDER BY last_visit_time DESC;
```



Ejecución de Filtro

Aplicar operador LIKE con wildcards para capturar todas las variantes de parámetros de búsqueda en diferentes motores.

Keyword Targeting

Buscar patrones específicos: search?q=, &query=, q=, adaptándose a Google, Bing, DuckDuckGo, etc.

Interpretación Contextual

Analizar la secuencia temporal de búsquedas para identificar cadenas de pensamiento, evolución de intereses o planificación de actividades.

Las búsquedas son la manifestación más directa del estado cognitivo del usuario en un momento dado. Una secuencia como "síntomas envenenamiento" → "comprar raticida online" → "borrar historial navegador" tiene valor probatorio evidente.

Correlación Temporal y Conclusiones del Perfilamiento

Análisis de Anomalías Temporales

Comparar el timestamp `last_visit_time` con horarios laborales estándar permite detectar patrones anómalos:

- **Actividad en madrugada (00:00-06:00):** Posible comportamiento furtivo
- **Fines de semana intensivos:** Actividad personal vs. laboral
- **Picos irregulares:** Eventos específicos que requieren investigación

Conclusión

El perfilamiento digital reduce exponencialmente el espacio de búsqueda forense.

En lugar de analizar gigabytes de datos sin dirección, priorizamos evidencia basada en la conducta detectada.



Estrategia de Investigación Dirigida

Si el perfilamiento detecta búsquedas como "borrar logs", "CCleaner", o "anti-forensics", el siguiente paso es buscar específicamente artefactos de herramientas de limpieza, registros de eventos borrados, o modificaciones en metadatos de archivos.

El perfilamiento no es el fin, es el mapa que guía la investigación hacia evidencia de alto valor probatorio.



Forense en Navegadores: Estructuras y Artefactos

Arquitectura Interna de Firefox (Motor Gecko)

Motor de Almacenamiento

Firefox utiliza SQLite como sistema de gestión de bases de datos para prácticamente todos sus componentes de almacenamiento persistente. Esta arquitectura unificada simplifica el análisis forense al proporcionar una estructura consistente y bien documentada.

Ubicación del Perfil

Los perfiles de usuario se almacenan en:

```
%AppData%\Roaming\Mozilla\Firefox\Profiles\<perfil_aleatorio>.default
```

Archivos Clave para Análisis

- **places.sqlite** - Historial de navegación y marcadores
- **cookies.sqlite** - Cookies y datos de sesión
- **formhistory.sqlite** - Datos de autocompletado de formularios



Arquitectura Chrome/Chromium



Motor Híbrido

Combina SQLite para datos estructurados y LevelDB para almacenamiento de alto rendimiento, ofreciendo flexibilidad en la persistencia de datos.



Ruta de Perfil

%LocalAppData%\Google\Chrome\User Data\Default\

Ubicación crítica para la adquisición forense de evidencia digital.



Archivos Esenciales

- **History** - Registro de URLs visitadas
- **Cookies** - Datos de sesión web
- **Login Data** - Credenciales almacenadas



Internet Explorer: Arquitectura Legacy

Internet Explorer representa un paradigma forense diferente debido a su antigüedad y estructura de almacenamiento basada en archivos planos. A pesar de su obsolescencia, sigue siendo relevante en investigaciones de sistemas Windows antiguos.

O1

Motor de Almacenamiento

Utiliza archivos planos y el formato propietario **index.dat** para indexación, un sistema menos eficiente pero recuperable mediante herramientas especializadas.

O2

Rutas Críticas

AppData\Local\Microsoft\Windows\History
Temporary Internet Files

O3

Modo InPrivate

Aunque diseñado para no guardar rastros en disco al cerrar, **deja evidencia recuperable en RAM y Pagefile** porque no realiza un borrado seguro (wipe) de la memoria.

Recuperación de Sesiones: Tab Recovery

Persistencia de Pestañas

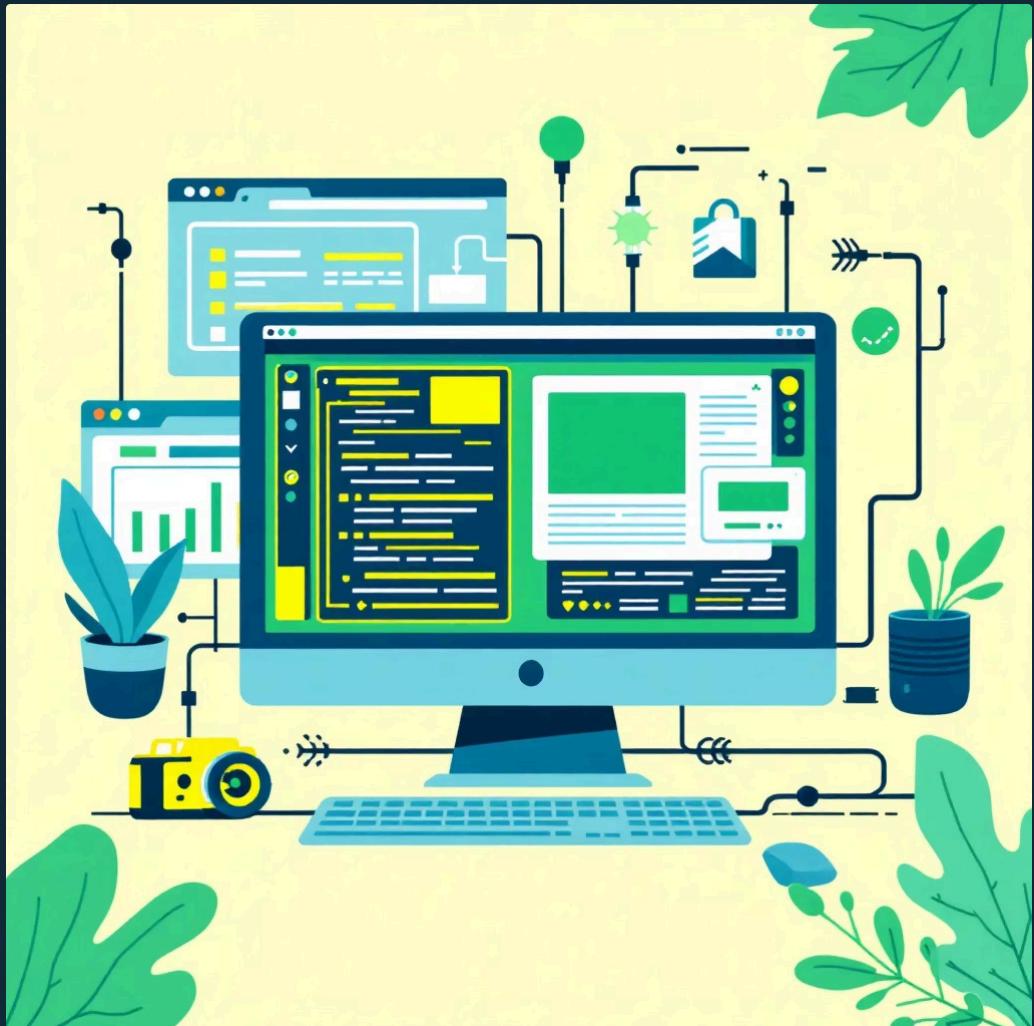
Los navegadores modernos mantienen registro de las pestañas abiertas para recuperarlas tras cierres inesperados o caídas del sistema. Este mecanismo forense es extremadamente valioso.

Ubicación en Internet Explorer

AppData\Local\Microsoft\Internet Explorer\Recovery\

Valor Forense

Contiene **URLs activas** que pueden no haberse registrado aún en el historial formal, proporcionando evidencia de actividad reciente inmediatamente antes del cierre.





Análisis de Caché y Artefactos Multimedia

El sistema de caché del navegador almacena recursos web localmente para optimizar tiempos de carga. Desde una perspectiva forense, representa una fuente rica de evidencia que incluye imágenes, scripts, hojas de estilo y contenido multimedia.

Propósito del Caché

Almacenamiento local de imágenes, scripts JavaScript, archivos CSS y otros recursos para acelerar la navegación subsecuente y reducir consumo de ancho de banda.

Herramientas Especializadas

ChromeCacheView - Para navegadores Chromium

MZCacheView - Para Firefox

Estas utilidades permiten visualizar y extraer archivos cacheados con sus metadatos asociados.

Ejemplo de Recuperación

Content-Type identificables:
image/jpeg, text/javascript,
text/css, video/mp4

Cada archivo mantiene información sobre su origen, fecha de descarga y tipo MIME.

Análisis Forense de Descargas

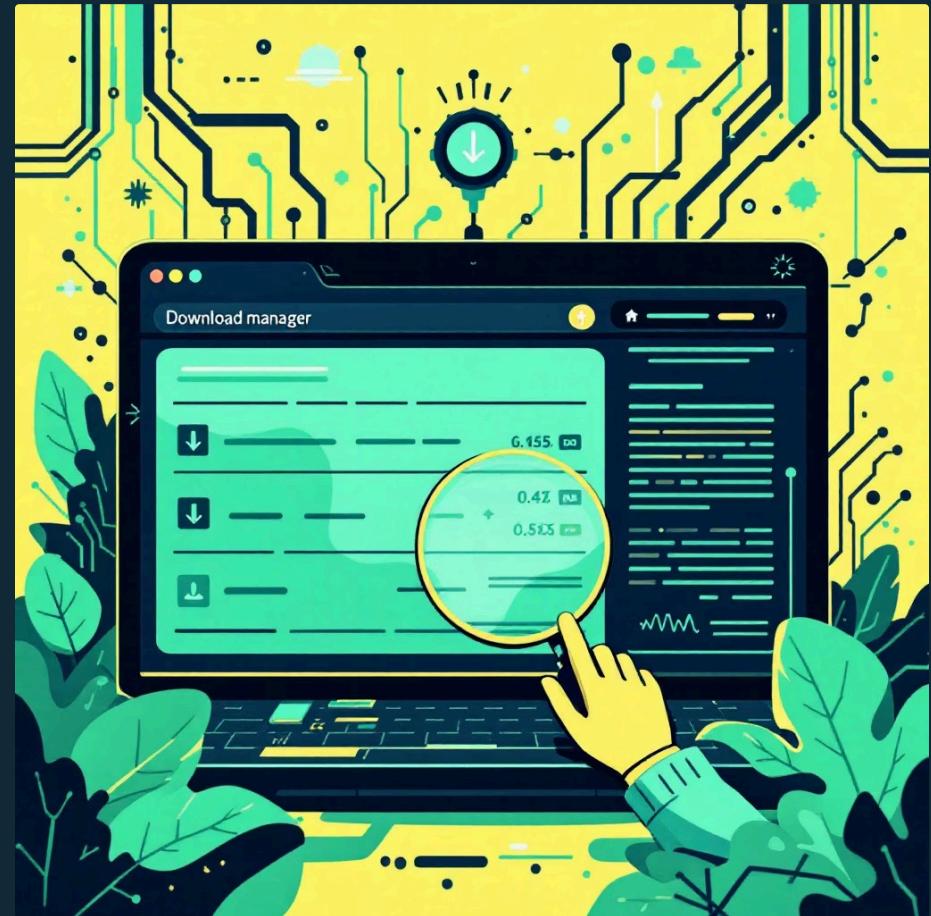
Distinción Crítica

Es fundamental diferenciar entre la **existencia de un registro de descarga** en la base de datos del navegador y la **presencia física del archivo** en el sistema de archivos. Un registro puede persistir incluso después de que el archivo haya sido eliminado.

Estructura de Datos

La tabla `downloads` en las bases SQLite contiene:

- Ruta de destino completa del archivo
- URL de origen de la descarga
- Referrer (página desde donde se inició)
- Timestamps de inicio y finalización
- Estado de la descarga (completada/interrumpida)



Suite de Herramientas NirSoft para Análisis Forense

Características Principales

Utilidades portables especializadas por navegador que no requieren instalación, facilitando su uso en entornos forenses controlados.

Lectura Directa

Acceso directo a estructuras de datos propietarias de navegadores sin necesidad de APIs intermedias, garantizando integridad de evidencia.

Consideración de Seguridad

Frecuentes falsos positivos en antivirus que las clasifican como "HackTools". Esto es esperable debido a sus capacidades de acceso profundo al sistema.



Demostración Práctica: Extracción y Validación



Extracción Automática

Utilización de **ChromeHistoryView** o **MZHistoryView** para análisis rápido de historial de navegación.

Resolución de problemas: La tecla **F9** permite apuntar a carpetas de perfil externas o perfiles corruptos que requieren análisis fuera de su ubicación original.



Validación Manual

Abrir `places.sqlite` en **DB Browser for SQLite** para inspección directa.

Navegar a la tabla `moz_places` y comparar datos crudos contra resultados parseados por herramientas automatizadas.



Verificación Cruzada

Contrastar múltiples fuentes de evidencia: historial, caché, descargas y cookies para construir una línea temporal coherente y verificable de la actividad del usuario.

Conclusiones: Diferencias Críticas en Almacenamiento

Datos "Roaming"

Configuración del usuario, historial de navegación, marcadores y preferencias que sincronizan entre dispositivos. Ubicación típica: %AppData%\Roaming

Valor forense: Persistente y sincronizable

Datos "Local"

Caché temporal, archivos de sesión y datos de rendimiento específicos de la máquina. Ubicación: %LocalAppData%

Valor forense: Volátil pero rico en artefactos recientes

- ☐ **Recomendación Crítica:** Siempre cerrar completamente el navegador antes de realizar la adquisición forense. Los navegadores modernos mantienen múltiples procesos en ejecución que pueden tener bases de datos bloqueadas, caché en memoria y datos sin persistir en disco, comprometiendo la integridad y completitud de la evidencia recolectada.



Forense en Emails: Headers y Rutas

Arquitectura y Flujo del Protocolo SMTP

Agentes de Correo

MUA (Mail User Agent): Cliente de correo que utiliza el usuario final para leer y enviar mensajes. Ejemplos incluyen Microsoft Outlook, Mozilla Thunderbird, Apple Mail y clientes web como Gmail.

MTA (Mail Transfer Agent): Servidor responsable de transferir correos entre sistemas. Ejemplos: Microsoft Exchange Server, Postfix, Sendmail, qmail.

Proceso de Transmisión

El flujo completo de un correo electrónico sigue esta secuencia: el emisor redacta el mensaje en su MUA, que lo envía al MTA propio del dominio. Este MTA lo transmite a través de Internet, pasando potencialmente por múltiples servidores intermedios, hasta llegar al MTA de destino, que finalmente lo deposita en el buzón del destinatario.



Formatos de Evidencia Digital en Correos

Formato .EML

Tipo: Archivo de texto plano codificado en ASCII

Ventaja: Estándar abierto, interoperable entre diferentes clientes de correo

Análisis: Puede abrirse directamente con editores de texto como Notepad++, facilitando la inspección forense sin herramientas especializadas

Compatibilidad: Thunderbird, Windows Mail, Outlook Express

Formato .MSG

Tipo: Archivo binario basado en estructura OLE (Object Linking and Embedding) de Microsoft

Desventaja: Formato propietario que requiere Microsoft Outlook o herramientas específicas para su análisis

Análisis: Necesita visualizadores hexadecimales o software especializado para extraer metadatos completos

Compatibilidad: Exclusivo del ecosistema Microsoft

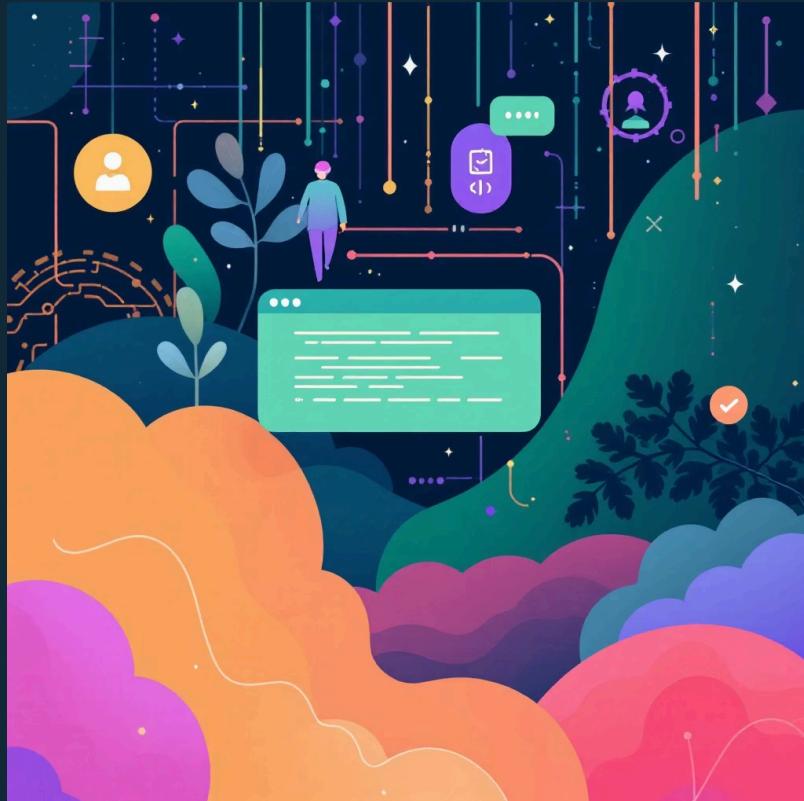
Almacenamiento Local

Outlook: .PST (Personal Storage Table) para archivos locales, .OST (Offline Storage Table) para caché sincronizada con servidor Exchange

Thunderbird: .MBOX, formato de texto que concatena múltiples correos en un único archivo

Consideración forense: Estos contenedores pueden almacenar miles de mensajes y requieren herramientas especializadas para su extracción

Cabeceras (Headers): La Metadata Reveladora



¿Qué son las cabeceras?

Las cabeceras son la parte invisible del correo electrónico que contiene información técnica sobre su transmisión. Se encuentran en la parte superior del código fuente del mensaje, ocultas para el usuario común pero accesibles mediante las opciones de visualización avanzada de cualquier cliente de correo.

Datos Volátiles Críticos

- **Dirección IP de origen:** Identifica el dispositivo o servidor desde el cual se envió originalmente el correo
- **Software emisor:** Información del cliente de correo utilizado (versión y tipo)
- **Ruta de servidores:** Secuencia completa de MTAs por los que transitó el mensaje
- **Marcas de tiempo:** Registro temporal en cada punto de la transmisión

Importante: Los campos **From** y **Reply-To** son fácilmente falsificables por el remitente. No deben utilizarse como única fuente de autenticación del origen.

Análisis de Ruta: El Campo "Received"



Campo Received

Cada servidor MTA por el que pasa un correo electrónico añade una línea "Received" al principio de las cabeceras. Estas líneas constituyen una traza completa y cronológica del recorrido del mensaje a través de la infraestructura de Internet.



Lectura Bottom-Up

La metodología correcta de análisis es de **abajo hacia arriba**. La última línea "Received" (la más inferior en el código) corresponde al primer servidor que procesó el mensaje, es decir, al origen real de la transmisión.



Línea Inferior = Origen Real

La línea "Received" inferior contiene la dirección IP real del remitente o del primer MTA que procesó el mensaje. Esta información es crítica para determinar la ubicación geográfica y el proveedor de servicios del emisor original.

Al examinar cada línea "Received", el investigador forense puede reconstruir la ruta exacta del correo, identificar demoras anormales entre saltos (que podrían indicar manipulación), y validar la coherencia del recorrido con los dominios involucrados.

Identificadores Únicos y Dirección de Retorno

Message-ID

String único generado automáticamente por el primer MTA que procesa el correo. Formato típico: <caracteres-aleatorios@dominio.com>

Uso forense: Permite realizar búsquedas precisas en logs de servidor (como /var/log/maillog en sistemas Unix o Exchange Tracking Logs en Windows) para confirmar la transmisión real del mensaje y detectar posibles duplicaciones o falsificaciones.



Return-Path

Dirección técnica utilizada por los servidores para enviar notificaciones de rebote (bounce messages) cuando un correo no puede ser entregado.

Indicador de Spoofing: Una discrepancia entre el campo "From" (visible para el usuario) y el "Return-Path" (técnico) es una señal clara de posible suplantación de identidad. El atacante puede mostrar una dirección legítima en "From" mientras que el "Return-Path" revela el dominio real controlado por él.

- En investigaciones forenses, siempre documente tanto el Message-ID como el Return-Path. Estos campos, junto con los "Received", forman la tríada de evidencia más confiable para establecer la autenticidad y trazabilidad de un correo electrónico.

Autenticación de Dominio: SPF, DKIM y DMARC

SPF (Sender Policy Framework)



Mecanismo que permite a los administradores de dominio publicar una lista de direcciones IP autorizadas para enviar correos en nombre de su dominio.

Funcionamiento: El servidor receptor consulta los registros DNS del dominio emisor para verificar si la IP de origen está autorizada.

Resultados:

- Pass: IP autorizada
- Fail: IP no autorizada
- SoftFail: Sospechoso
- Neutral: Sin política

DKIM (DomainKeys Identified Mail)



Firma criptográfica digital que el servidor emisor añade al correo, garantizando que el contenido no ha sido alterado durante la transmisión.

Funcionamiento: El servidor firma el mensaje con una clave privada. El receptor verifica la firma usando la clave pública publicada en DNS.

Ventaja forense: Permite detectar cualquier modificación del contenido o las cabeceras después del envío original.

Interpretación Forense



Pass (Legítimo): Tanto SPF como DKIM validan correctamente. Alta probabilidad de autenticidad.

Fail (Spoofing): Fallo en validación indica posible suplantación de identidad o correo no autorizado.

SoftFail: Señal de alerta que requiere análisis adicional de otros indicadores.

None: Dominio sin políticas configuradas (común en dominios pequeños o personales).

DEMO PRÁCTICA 1: Análisis Raw con Notepad++

O1

Preparación del Archivo

Obtener el correo en formato .EML. Si el correo está en Outlook, exportarlo como .EML o usar "Ver código fuente" y copiar todo el contenido a un archivo de texto.

O2

Apertura en Notepad++

Abrir el archivo .EML con Notepad++ (o cualquier editor de texto avanzado con búsqueda de expresiones regulares y numeración de líneas).

O3

Localizar Campo Received Inferior

Usar la función de búsqueda (Ctrl+F) para encontrar todas las ocurrencias de "Received:". Desplazarse hasta la última aparición (más cercana al final de las cabeceras). Esta línea contiene la IP de origen real.

O4

Identificar Software Emisor

Buscar los campos "X-Mailer:" o "User-Agent:". Estos revelan el cliente de correo utilizado y su versión, información valiosa para perfilamiento técnico del emisor.

O5

Extraer Direcciones IP

Anotar todas las IPs encontradas en los campos "Received" para posterior geolocalización y verificación en bases de datos de reputación de IP.

Ventaja del análisis raw: Permite ver exactamente la estructura original del correo sin interpretaciones de software, garantizando que no se pierda información crítica por procesamiento automático.

DEMO PRÁCTICA 2: Visualización con Google Messageheader



Procedimiento

1. Copiar todas las cabeceras del correo (desde el principio hasta la línea en blanco que precede al cuerpo del mensaje)
2. Pegar el contenido en el campo de texto de la herramienta Google Messageheader
3. Hacer clic en "Analyze header"

Resultados Obtenidos

- **Visualización de "Hops":** Gráfico que muestra cada salto entre servidores
- **Tiempo de demora (Delay):** Análisis temporal entre cada hop, útil para detectar demoras anormales
- **Validación SPF/DKIM:** Resultados de autenticación claramente indicados con iconos de verificación o advertencia

Herramienta Utilizada

Google Messageheader Analyzer es una herramienta web gratuita que proporciona visualización gráfica del recorrido de un correo electrónico.

URL: toolbox.googleapps.com/apps/messageheader/

Geolocalización Manual

Con la IP de origen identificada, utilizar servicios como MaxMind GeolP, IPinfo.io, o bases de datos WHOIS para determinar la ubicación geográfica, el proveedor de servicios (ISP), y la organización propietaria del rango de IPs. Esta información contextualiza el origen del correo y puede revelar inconsistencias con la identidad declarada del remitente.

Conclusión: Las Cabeceras No Mienten

Principio Fundamental

El cuerpo del correo y los campos visibles (From, Subject, Reply-To) pueden ser fácilmente manipulados por un atacante. Sin embargo, las cabeceras técnicas, especialmente los campos "Received", "Message-ID" y los resultados de SPF/DKIM, proporcionan evidencia forense mucho más confiable sobre el verdadero origen y ruta del mensaje.

Preservación de Evidencia

Es absolutamente crítico preservar el formato original del correo (.eml o .msg) en cualquier investigación forense. Reenviar un correo o hacer capturas de pantalla destruye las cabeceras originales y hace imposible la verificación técnica del origen. La cadena de custodia digital debe incluir el archivo completo sin modificaciones.

Metodología de Análisis

Un análisis forense completo de correo electrónico debe combinar: (1) Inspección manual raw de cabeceras, (2) Validación de autenticación SPF/DKIM, (3) Trazado de ruta mediante campos "Received", (4) Geolocalización de IPs de origen, y (5) Correlación con logs de servidor cuando sea posible. Solo esta aproximación integral garantiza conclusiones sólidas.

- **Recordatorio final:** En el análisis forense de correos electrónicos, la metadata es más valiosa que el contenido visible. Las cabeceras revelan la verdad técnica que los atacantes no pueden ocultar completamente.