MF0952_2: Publicación de páginas web

1. Características de seguridad en la publicación de páginas web

Los sistemas de Archivos

La mayoría de los sistemas operativos poseen su propio sistema de archivos. Los sistemas de archivos son representados ya sea textual o gráficamente utilizando gestores de archivos o “shells”. En modo gráfico a menudo son utilizadas las metáforas de carpetas (directorios) conteniendo documentos, archivos y otras carpetas. Un sistema de archivos es parte integral de un sistema operativo moderno.

Los sistemas de archivos más comunes utilizan dispositivos de almacenamiento de datos que permiten el acceso a los datos como una cadena de bloques de un mismo tamaño, a veces llamados sectores, usualmente de 512 bytes de longitud. El software del sistema de archivos es responsable de la organización de estos sectores en archivos y directorios y mantiene un registro de qué sectores pertenecen a qué archivos y cuáles no han sido utilizados. En la realidad, un sistema de archivos no requiere necesariamente de un dispositivo de almacenamiento de datos, sino que puede ser utilizado también para acceder a datos generados dinámicamente, como los recibidos a través de una conexión de red.

Generalmente un sistema de archivos tiene directorios que asocian nombres de archivos con archivos, usualmente conectando el nombre de archivo a un índice en una tabla de asignación archivos de algún tipo, como FAT en sistemas de archivos MS-DOS o los inodos de los sistemas Unix.

La estructura de directorios puede ser plana o jerárquica (ramificada o “en árbol”). En algunos sistemas de archivos los nombres de archivos son estructurados, con sintaxis especiales para extensiones de archivos y números de versión. En otros, los nombres de archivos son simplemente cadenas de texto y los metadatos de cada archivo son alojados separadamente.

En sistemas de archivos jerárquicos, en lo usual, se declara la ubicación precisa de un archivo con una cadena de texto llamada “ruta”. La nomenclatura para rutas varía ligeramente de sistema en sistema, pero mantienen por lo general una misma estructura. Una ruta viene dada por una sucesión de nombres de directorios y subdirectorios, ordenados jerárquicamente de izquierda a derecha y separados por algún caracter especial que suele ser una barra (‘/’) o barra invertida (‘\’) y puede terminar en el nombre de un archivo presente en la última rama de directorios especificada.

Así, por ejemplo, en un sistema Unix la ruta al archivo deseado del usuario “user” sería algo como:

/home/user/documentos personales/archivo.doc

Un ejemplo análogo en un sistema de archivos Windows se vería como:

C:\Mis documentos\documentos personales/archivo.doc

Los sistemas de archivos pueden ser clasificados en tres ramas:

Sistemas de archivos de disco.
Sistemas de archivos de red.
Sistemas de archivos de propósito especial.

Sistemas de archivos de disco

Un sistema de archivo de disco está diseñado para el almacenamiento de archivos en una unidad de disco, que puede estar conectada directa o indirectamente a la computadora.

Tipos:

Ext2 y 3.
FAT16 y 32.
NTFS.
ReiserFS
ISO 9660.

EXT2 (Second extended Filesystem o "Segundo sistema de archivos extendido"):

Fue el sistema de archivos estándar en el sistema operativo Linux por varios años y continúa siendo ampliamente utilizado.

Fue diseñado originalmente por Rémy Card.

La principal desventaja de EXT2 es que no posee una bitácora, por lo que muchos de sus usuarios están emigrando a ReiserFS y su sucesor EXT3.

EXT3 (Third extended Filesystem o "Tercer sistema de archivos extendido"):

Es un sistema de archivos con registro por diario (en inglés "journaling"), el cual se encuentra creciendo en popularidad entre usuarios del sistema operativo Linux.

A pesar de su menor desempeño y escalabilidad frente a alternativas como ReiserFS o XFS, posee la ventaja de permitir migrar del sistema de archivos EXT2 sin necesidad de reformatear el disco.

La única diferencia entre EXT2 y EXT3 es el registro por diario.

Un sistema de archivos EXT3 puede ser montado y usado como un sistema de archivos EXT2.

FAT (File Allocation Table o "Tabla de ubicación de archivos"):

Es el principal sistema de archivos desarrollado para MS-DOS y Windows.

El sistema de archivos FAT es relativamente sencillo, y debido a eso es muy popular como formato para disquetes. Además, el formato FAT es soportado por casi todos los sistemas operativos para PCs IBM, y debido a esto a menudo se lo utiliza para compartir información entre diversos sistemas operativos en un mismo equipo.

FAT es un sistema de archivos relativamente anticuado, y debido a esto sufre de varios problemas:

Para comenzar, su distribución de archivos simple permite la fragmentación, lo que produce eventuales pérdidas en el desempeño de operaciones sobre archivos.

Luego, FAT no fue diseñado para redundancia en caso de fallas del sistema.

Las primeras versiones de FAT permitían nombres de archivo de hasta 12 caracteres, aunque esto fue solucionado por Microsoft al inventar VFAT, el cual permite nombres de hasta 255 caracteres.

Finalmente, los sistemas de archivos FAT no permiten directivas de seguridad, garantizando el acceso a todos los archivos de una partición por cualquier usuario del sistema operativo.

NTFS (New Tecnology File System o "Sistema de archivos de nueva tecnología"):

Sistema de archivos diseñado específicamente para Windows NT, con el objetivo de crear un

sistema de archivos eficiente, robusto y con seguridad incorporada desde su base.

También soporta compresión nativa de ficheros y encriptación (esto último sólo a partir de Windows 2000).

NTFS permite definir el tamaño del cluster, a partir de 512 bytes (tamaño mínimo de un sector) de forma independiente al tamaño de la partición.

Es un sistema adecuado para las particiones de gran tamaño requeridas en estaciones de trabajo de alto rendimiento y servidores.

Puede manejar discos de hasta 2 terabytes.

Los inconvenientes que plantea son:

Necesita para si mismo una buena cantidad de espacio en disco duro por lo que no es recomendable su uso en discos menores de 400 MB.

No es compatible con MS-DOS, Windows 95 ni Windows 98.

La conversión a NTFS es unidireccional, si elige actualizar la unidad, no podrá volver a convertirla a FAT.

ReiserFS:

Es un sistema de archivos de propósitos generales, diseñado e implementado por un equipo liderado por Hans Reiser.

Actualmente funciona bajo Linux, con la versión 2.4.1 del núcleo Linux, se convirtió en el primer sistema de archivos con registro por diario (en inglés, “journaling”) en ser incluido en el núcleo estándar.

La ventaja más evidente sobre el sistema de archivos estándar de Linux, EXT2, es su registro por diario.

Esto reduce ampliamente el riesgo de corrupción del sistema de archivos (y la necesidad de extensas revisiones del sistema) después de un apagado no programado del sistema, ya sea por un corte eléctrico o un error del sistema.

Desafortunadamente, convertir un sistema a ReiserFS requiere para usuarios de EXT2 el reformateo completo de sus discos, una desventaja no presente en su principal competidor, EXT3.

ReiserFS maneja directorios conteniendo enormes cantidades de archivos pequeños muy eficientemente.

ISO 9660:

En el año 1985, diferentes distribuidores de software y fabricantes de hardware trabajaron conjuntamente obteniendo como fruto el llamado formato HSG, vigente aún hoy en día en los CD para ordenadores PC y también para muchos sistemas UNIX.

Todos los CD-ROM actuales están provistos de este formato.

El nombre de este formato viene de “High Sierra Group”, que es el nombre que recibieron los diferentes técnicos que participaron en el desarrollo del HSG en honor al primer lugar donde se reunieron, el hotel y casino “High Sierra” en el estado de Nevada, Estados Unidos.

Un año después, las autoridades de normalización americanas ISO estandarizaron la propuesta, que se presentó bajo el título “Volume and File Structure of Compact Read Only Optical Disk for Information Interchange”.

Desde entonces se habla de la norma ISO 9660 o simplemente de la ISO 9660.

1.1 Seguridad en los distintos tipos de archivos

Una norma básica de seguridad radica en la asignación a cada usuario sólo de los permisos necesarios para poder cubrir las necesidades de su trabajo sin poner en riesgo el trabajo de los demás.

INTRODUCCION

En la actualidad, cada vez más aplicaciones y servicios se desarrollan para entornos web -siendo accesibles a través del navegador- de forma que se facilite el acceso desde las redes internas de las empresas o a través de Internet. Por este motivo, se debe poner especial énfasis en la seguridad de los servidores web.

Por poner un ejemplo, caídas del servidor, problemas se seguridad por la configuración, robo de información confidencial, defacements (modificaciones en del aspecto de la web) o inyección de código malicioso son algunas de las consecuencias de no disponer una instalación segura del servidor web.

Debido a la facilidad de instalación, las extensas posibilidades de configuración y los módulos de seguridad que dispone, Apache se ha convertido en el servidor web más utilizado entre el resto de opciones. (Instalación de Apache)

Los temas de seguridad se puede analizar desde dos perspectivas diferentes, la seguridad externa y la interna. En un sistema informático, con el fin de asegurar la integridad de la información contenida en él, se precisa describir las directrices necesarias y los mecanismos capaces de satisfacerlas para lograr dicho fin. Dependiendo de los mecanismos utilizados y de su grado de efectividad, se puede hablar de sistemas seguros e inseguros.

En primer lugar, deben imponerse ciertas características en el entorno donde se encuentra la instalación de los equipos, con el fin de impedir el acceso a personas no autorizadas, mantener un buen estado y uso todo el material y equipos, así como eliminar los riesgos de causas de fuerza mayor (por ejemplo incendios, inundaciones, etc.) que puedan destruir la instalación y la información contenida.

En la actualidad son muchas las violaciones que se producen en los sistemas informáticos, en general por accesos de personas no autorizadas que obtienen información confidencial, pudiendo incluso manipularla. En ocasiones, este tipo de incidencias resulta grave por la naturaleza de los datos; por ejemplo, si se trata de datos bancarios, datos oficiales que pueden afectar a la seguridad de los Estados, etc.

Ahora bien, no todas las violaciones se deben a accesos no permitidos, sino que pueden producirse por muy diversas causas. Entre éstas, una que últimamente se ha puesto de moda con el progreso de la microinformática es el software malintencionado, es decir, pequeños programas que poseen una gran facilidad para reproducirse y ejecutarse, cuyos efectos son destructivos y el daño en la mayoría de los casos es irreversible. Nos estamos refiriendo a lo que en términos populares se ha dado en llamar virus informático.

Por todo ello, los gobiernos de los distintos países han dictado leyes y normas para asegurar una racional seguridad en los sistemas de información y proteger el derecho a la intimidad de la información de las personas.

Por último, un aspecto que no debemos olvidar es el de evaluar el nivel de seguridad que una instalación necesita. Dependerá fundamentalmente de la importancia de los datos, su grado de confidencialidad, etc.; por tanto, no será igual la seguridad que necesita una instalación bancaria que un equipo de uso doméstico.

SEGURIDAD EXTERNA

En un sistema informático todos los mecanismos de seguridad tienen que complementarse entre sí, de tal forma que si una persona logra saltarse alguna de las protecciones, se encuentre con otras que le hagan el camino difícil.

Todos los mecanismos dirigidos a asegurar el sistema informático sin que el propio sistema intervenga en el mismo se engloban en lo que podemos denominar seguridad externa.

La seguridad externa puede dividirse en dos grandes grupos:

Seguridad física. Engloba aquellos mecanismos que impiden a los agentes físicos la destrucción de la información existente en el sistema; entre ellos podemos citar el fuego, humo, inundaciones, descargas eléctricas, campos magnéticos, acceso físico de personas con no muy buena intención, etc.

Seguridad de administración. Engloba los mecanismos más usuales para impedir el acceso lógico de personas físicas al sistema.

Seguridad física

Como ya hemos mencionado, se trata de eliminar los posibles peligros que originan los agentes físicos o la presencia física de personas no autorizadas. Para ello podemos considerar los siguientes aspectos:

1. **Protección contra desastres.** Consta de elementos de prevención, detección y eliminación que actúan contra incendios, humos, sobretensiones, fallos en el suministro de energía, etc. También es necesario controlar la temperatura y limpieza del medio ambiente en que se encuentran los equipos, instalando aire acondicionado, falso suelo, ventilación, y, en definitiva, tomando en consideración todo aquello que pueda causar cualquier problema a la instalación.
2. **Protección contra intrusos.** Desde el punto de vista físico, es necesario establecer mecanismos que impidan el acceso físico de las personas no autorizadas a las instalaciones. Suele llevarse a cabo mediante puertas de seguridad con apertura por clave o llaves especiales, identificación de las personas por tarjetas de acceso o por reconocimiento de la voz, huellas digitales, etc.

Seguridad de administración

Comprende aquellos mecanismos cuya misión es dar acceso lógico al sistema. Este acceso puede realizarse a través de un terminal del sistema o bien desde otro sistema por medio de una red de comunicación a la que estén conectados ambos sistemas.

Protección de acceso

Se trata de un mecanismo para el control de los intentos de entrada o acceso al sistema, de tal forma que permita la conexión cuando un usuario lo solicite y pase el control correspondiente y rechace el intento en aquellos casos en que la identificación del supuesto usuario no sea satisfactoria.

Palabra de acceso o identificador del usuario (password). Para la identificación del usuario, la fórmula más extendida es la de pedirle su nombre de usuario (username) y a continuación la palabra clave tal que el mecanismo accede al archivo correspondiente para contrastar los datos recibidos y aceptar o rechazar el intento. Los intentos fallidos de acceso son registrados por el sistema con el fin de que el administrador del sistema pueda estudiar cada cierto tiempo si se está o no intentando transgredir la seguridad del sistema.

El sistema operativo dota al administrador del sistema para que en cualquier momento se pueda dar de alta o de baja a un usuario, asignándole en el primer caso, además de un username, la correspondiente contraseña o password inicial. Mientras que el nombre de usuario es público, el

password no lo es, siendo recomendable su cambio cada cierto tiempo, así como no tenerla escrita en ninguna otra parte que en la propia mente del usuario.

El password cuando se escribe en un terminal, tanto para acceder al sistema como para su cambio, no aparece en la pantalla como ocurre en el resto de datos que se teclean, para así conservar el secreto de la misma. Además, esta palabra se graba en los archivos de administración del sistema codificada o encriptada para que no sea fácilmente reconocible por las personas.
Al proceso de petición de entrada a un sistema, contestación a las preguntas de identificación, contrastación de los datos recibidos y dar el correspondiente acceso se denomina login . Asimismo, al proceso de despedida del sistema se le llama logout.

Criptografia. Es un proceso de transformación que se aplica a unos datos para ocultar su contenido. El proceso al que hay que someter la información para conseguir que sea secreta se conoce con el nombre de encriptado o cifrado, denominándose la información antes del proceso como texto claro y después del mismo texto cifrado.

La mayoría de los sistemas que utilizan algoritmos de cifrado exigen que el texto cifrado pueda convertirse en texto claro. Para ello existen diversas técnicas, algunas de las cuales describimos brevemente a continuación:

Or-exclusivo. Es un método sencillo que ofrece una gran seguridad. Consiste en tomar la información a cifrar y aplicar a cada octeto la operación or-exclusivo con una clave cuya longitud debe ser, en número de caracteres, tan larga como el mensaje a cifrar. El algoritmo de descifrado es similar al de cifrado utilizando la misma clave. Esta clave se tiene que cambiar cada cierto tiempo para mantener un buen grado de seguridad.

Estándar de Encriptado de Datos (Data Encryption Standard-DES). Es un método desarrollado en la Oficina Nacional de Estándares de Estados Unidos, siendo uno de los más utilizados actualmente. El algoritmo lleva asociado un chip especial-mente construido para este fin, aunque puede ser simulado por software, y se basa en claves de 56 bits de longitud.
Método de Rivest, Shamir y Adelman (RSA). Es un algoritmo que utiliza distinta clave para el cifrado y descifrado, ofreciendo con ello un alto índice de seguridad.

Seguridad funcional
Engloba aspectos relativos al funcionamiento del sistema y a la seguridad que de las instalaciones se pretende tener.
· Seguridad en la transmisión de datos. En las líneas de transmisión de datos existen diversos problemas de seguridad debido a lo fácilmente violables que son dichas líneas. Por esta razón, para enviar datos a través de líneas de comunicación entre computadoras se siguen diversas técnicas, como son:

a) Compactación de datos. Consiste en comprimir los datos para que ocupen el menor espacio posible y así conseguir en principio que la duración de la transmisión sea menor, y que para entenderla haya que descompactarla; por tanto, la información va relativamente cifrada. Existen muchos métodos de compactación de datos, de los cuales los más utilizados son:

1. Reducción de espacios en blanco. Un archivo de información puede tener muchos espacios en blanco que pueden ser sustituidos por un número que indique cuántos de ellos están de forma consecutiva en un determinado punto.
2. Codificación por diferencia. En ella se transmiten únicamente las diferencias existentes entre la información que se quiere enviar y la misma información ya enviada previamente, de tal forma que en el destino se puede reconstruir la información sin grandes dificultades. Se trata de un caso

similar a las copias de seguridad (Backup) incrementales, donde cada nueva copia sólo registra las diferencias que existan entre el estado actual de la información y el original, con lo que se logra un importante ahorro de memoria.

b) Criptografía. Similar al proceso ya mencionado para ocultar la información en una transmisión.
c) Fiabilidad. Además de las medidas anteriores, se suelen tomar otras para asegurar el correcto estado de la información al llegar a su destino. Se pueden presentar problemas debidos a causas accidentales, como la influencia de fuertes campos magnéticos, perturbaciones eléctricas, etc., así como por motivos de intrusión en las comunicaciones con el fin de destruirlas o modificarlas. También pueden producirse errores por colisiones entre mensajes en redes locales y un sinfín de otras causas de diversa naturaleza.

Para evitar todo tipo de incidencias, se suele añadir a la información una pequeña parte que nos permitirá saber silos datos recibidos coinciden con los enviados o no. Los métodos más utilizados para dotar de fiabilidad a una transmisión de datos son mecanismos hardware o software que permiten detectar errores ocurridos en una comunicación e incluso recuperar algunos de ellos. Citaremos los siguientes métodos:

1. Bit de paridad. Consiste en añadir un bit a cada octeto o palabra que se transmita para con él conseguir que la suma de unos sea par (paridad par) o impar (paridad impar). Con este método se detectan errores al variar un bit o un numero impar de ellos sin que se detecten variaciones de un número par de bits. Se sabe que la mayoría de errores que se producen en condiciones normales sólo afectan a un bit.
2. Códigos de Hamming. Añaden varios bits de control al octeto o palabra a transmitir, de tal forma que detectan errores de uno o más bits y los corrigen.

3. Código de redundancia cíclica (CRC). Si se prevé que los daños esperados en una transmisión no sean de un bit en un octeto o palabra, sino en una secuencia de ellos, se puede utilizar un algoritmo que permita realizar una suma denominada suma de chequeo (Checksum) y aplicar el método denominado de redundancia cíclica durante la transmisión, de tal forma que al terminar ésta se repite en el destino el mismo algoritmo de suma, comprobándose si el valor final de la suma es el mismo.

· Sistemas tolerantes a fallos. Se utilizan en sistemas donde se pueda perder información debido a un mal funcionamiento de los mismos. Este aspecto es muy importante en los sistemas de control y supervisión en tiempo real. Existen mecanismos que ante situaciones de mal funcionamiento consiguen recuperar y controlar el entorno, protegiendo fundamentalmente la información. Este tipo de mecanismos se basa en redes de dos o más computadoras conectadas entre sí de manera que, ante el mal funcionamiento de una de ellas, éste se pondrá en situación de inactivo, tomando el control cualquiera de los otros que estén conectados.

SEGURIDAD INTERNA

Todos los mecanismos dirigidos a asegurar el sistema informático, siendo el propio sistema el que controla dichos mecanismos, se engloban en lo que podemos denominar seguridad interna.

Seguridad del procesador

Los mecanismos de protección del procesador son varios ya estudiados y que pasamos a enumerar:

Estados protegidos (Kernel) o no protegido (Usuario).

Reloj hardware para evitar el bloqueo del procesador.

Seguridad de la memoria

Se trata de mecanismos para evitar que un usuario acceda la información de otro sin autorización. Entre ellos citaremos dos:

- Registros limites o frontera.
- Estado protegido y no protegido del procesador.

Además se emplean para la memoria métodos como el de utilizar un bit de paridad o el checksum ya mencionado.

Seguridad de los archivos

La finalidad principal de las computadoras es la del tratamiento de la información que se almacena permanentemente en los archivos. La pérdida o alteración no deseada de dicha información causaría trastornos que podrían ser irreparables en algunos casos. Por eso es necesario tomar las correspondientes medidas de seguridad, que se deben enfocar desde dos aspectos diferentes: la disponibilidad y la privacidad de los archivos.

Disponibilidad de los archivos
Un archivo debe tener la información prevista y estar disponible en el momento que un usuario la necesite. Hay que tener presente la necesidad de asegurar tal circunstancia y para ello se suelen realizar las siguientes acciones:

Copias de seguridad (backup).
Consiste en que cada cierto tiempo (hora, día, semana…) se realice una copia del contenido de los archivos, de forma que si se destruyen éstos, es posible la recuperación de los datos a partir de la última de las copias. La operación de realizar copias de seguridad, así como la recuperación de los datos a partir de las mismas, se suele hacer por medio de programas de utilidad del sistema operativo.
La fiabilidad de las copias de seguridad dependerá fundamentalmente de la periodicidad con que se realicen y del índice de actividad de los archivos, es decir, del ritmo al que se actualicen.
Las copias de seguridad suelen realizarse sobre cinta magnética, guardándose en dependencias alejadas del sistema y en armarios protegidos incluso contra incendios.
Al margen de las copias de seguridad, en muchos casos es conveniente mantener los archivos duplicados en el mismo o distinto disco, para que en caso de problemas locales en el archivo original se pueda tener una rápida recuperación. En los grandes sistemas se tiende a automatizar los procesos de copias de seguridad por medio de un software que periódicamente revisa la fecha de la última copia de cada archivo, así como su último proceso de actualización, y a través de unos parámetros prefijados decide en qué archivos deben ser procesadas sus copias.

Archivos LOG.
En sistemas de tiempo compartido donde trabajan simultáneamente muchos usuarios, que entre otras operaciones llevan a cabo numerosas actualizaciones y modificaciones de archivos, no son suficientes las periódicas copias de seguridad para afrontar la pérdida de la información. Si la computadora falla por cualquier motivo en medio de una sesión donde hay un gran número de usuarios trabajando, se puede recuperar la información de los archivos desde la última copia de seguridad; pero esto puede no ser suficiente, por lo cual se recurre en estos sistemas a archivos auxiliares donde se registran todas las operaciones que realiza un usuario sobre sus archivos, almacenándose la nueva información o aquella que difiera de la ya existente. Estos archivos reciben

el nombre de archivos LOG y son tratados por utilidades del sistema operativo conjuntamente con las copias de seguridad para los procesos de recuperación.

Privacidad de los archivos
El contenido de los archivos se debe proteger de posibles accesos no deseados. Entre el peligro de permitir a todos los usuarios el acceso a cualquier archivo, y la rigidez de que cada usuario sólo pueda acceder a los suyos, el sistema de protección debe permitir accesos de forma controlada, según unas reglas predefinidas y con las consiguientes autorizaciones.
Cada usuario, al comenzar la sesión en un sistema tras su identificación, tiene asignado por el sistema de protección un dominio compuesto de una serie de recursos y de operaciones permitidas, por ejemplo, una serie de archivos a los que acceder, no teniendo permitido el acceso al resto de archivos. En general, los sistemas operativos almacenan la información relativa a los dominios en lo que se denomina matriz de dominios, cuyas filas indican los dominios existentes y las columnas los recursos. Cada elemento de la matriz indica el derecho a utilizar el recurso correspondiente en el dominio.

Si la matriz anterior tiene poca información, se recurre a otro tipo de almacenamiento de información sobre dominios, consistente en asociar a cada recurso una lista de dominios que pueden utilizarlo, denominándose éste vector lista de acceso. También se puede obtener otro vector donde a cada dominio se le asigna una lista de recursos a los que puede acceder, denominándose en este caso lista de capacidades.

En todos estos casos, la gestión de las listas de control se realiza mediante comandos de uso restringido, estando éstos únicamente disponibles para el administrador del sistema.

1.2 Permisos de acceso

Linux, como sistema multiusuario, asigna un propietario y un grupo a cada fichero (y directorio) y unos permisos al propietario, al grupo y al resto de los usuarios. La forma más rápida de comprobar esta característica es usar el comando ls -la. Así nos aparece el tipo de fichero, el propietario, el grupo, los permisos e información adicional. Por supuesto, el sistema de ficheros tiene que admitir esta característica, como es el caso del sistema de ficheros ext2 (Linux nativo). En los sistemas de ficheros pensados para entornos monousario, como msdos o vfat, no tenemos esta característica, por lo que son inseguros y su uso no es aconsejable bajo Linux.

Es conveniente tener claros los permisos que se pueden asignar a un fichero o directorio. Puede que algunas aplicaciones no funcionen bien si algún fichero no tiene el permiso o el propietario correctos, bien por falta de permisos o bien por exceso. Algunas aplicaciones son un poco quisquillosas en este aspecto. Por ejemplo, fetchmail es un programa que podemos usar para recoger el correo de un servidor pop (por ejemplo). Este programa se configura en el fichero .fetchmailrc, donde tendremos que indicar la clave que usamos en el servidor; pues bien, si este fichero tiene permiso de lectura para otro usuario que no sea el propietario, fetchmail no funcionará.

Otras aplicaciones, como por ejemplo inn (un servidor de noticias de Internet) tampoco funcionará si el propietario de sus ficheros es otro usuario distinto a news. Todo esto está perfectamente documentado en cada uno de los programas, por lo que es conveniente leer la documentación que aporta y las páginas del manual.

Permisos de ficheros y directorios

Como decíamos anteriormente, tenemos que asegurarnos de que los ficheros del sistema y los de

cada usuario sólo son accesibles por quienes tienen que hacerlo y de la forma que deben. No sólo hay que protegerse de ataques o miradas indiscretas, también hay que protegerse de acciones accidentales.

En general, cualquier sistema UNIX divide el control de acceso a ficheros y directorios en tres elementos: propietario, grupo y otros. Tanto el propietario como el grupo són únicos para cada fichero o directorio. Eso sí, a un grupo pueden pertenecer múltiples usuarios. Otros hace referencia a los usuarios que ni son el propietario ni pertenecen al grupo.

Todas estas características se almacenan en el sistema de ficheros, en particular en un i-nodo, que es un elemento que describe las características de un fichero en disco (salvo su nombre).

Una rápida explicación de los permisos Unix:

Propiedad:
Qué usuario y grupo posee el control de los permisos del i-nodo. Se almacenan como dos valores numéricos, el uid (user id) y gid (group id).
Permisos:
Bits individuales que definen el acceso a un fichero o directorio. Los permisos para directorio tienen un sentido diferente a los permisos para ficheros. Más abajo se explican algunas diferencias.
Lectura (r):
Fichero: Poder acceder a los contenidos de un fichero
Directorio: Poder leer un directorio, ver qué ficheros contiene
Escritura (w):
Fichero: Poder modificar o añadir contenido a un fichero
Directorio: Poder borrar o mover ficheros en un directorio
Ejecución(x):
Fichero: Poder ejecutar un programa binario o guion de shell
Directorio: Poder entrar en un directorio
Estos permisos se pueden aplicar a:

usuario (u): El propietario del fichero
grupo (g): El grupo al que pertenece el fichero
otros (o): El resto de los usuarios del sistema
Nota:
Tenga mucho cuidado con aquellos directorios que tengan permiso de escritura. Cualquiera podría borrar un fichero, aunque no sea de su propiedad y esto puede ser un riesgo, tanto para el sistema como para los datos de los usuarios.
Además tenemos otros bits de permisos que no podemos pasar por alto cuando estamos tratando de temas de seguridad.

Sticky bit:
El sticky bit tiene su significado propio cuando se aplica a directorios. Si el sticky bit está activo en un directorio, entonces un usuario sólo puede borrar ficheros que son de su propiedad o para los que tiene permiso explícito de escritura, incluso cuando tiene acceso de escritura al directorio. Esto está pensado para directorios como /tmp, que tienen permiso de escritura global, pero no es deseable permitir a cualquier usuario borrar los ficheros que quiera. El sticky bit aparece como ‘t’ en los listados largos de directorios.
drwxrwxrwt 19 root root 8192 Jun 24 14:40 tmp

Attributo SUID: (Para Ficheros)
Este bit describe permisos al identificador de usuario del fichero. Cuando el modo de acceso de

ID de usuario está activo en los permisos del propietario, y ese fichero es ejecutable, los procesos que lo ejecutan obtienen acceso a los recursos del sistema basados en el usuario que crea el proceso (no el usuario que lo lanza). Por ejemplo /usr/bin/passwd es un ejecutable propiedad de root y con el bit SUID activo. ¿Por qué? Este programa lo puede usar cualquier usuario para modificar su clave de acceso, que se almacena en
-rw-r–r– 1 root root 1265 Jun 22 17:35 /etc/passwd

pero según los permisos que observamos en este fichero, sólo root puede escribir y modificar en él. Entonces sería imposible que un usuario pudiera cambiar su clave si no puede modificar este fichero. La solución para este problema consiste en activar el bit SUID para este fichero:

-r-s–x–x 1 root root 10704 Apr 14 23:21 /usr/bin/passwd

de forma que cuando se ejecute, el proceso generado por él es un proceso propiedad de root con todos los privilegios que ello acarrea.

¿Piensa que esto puede ser un riesgo para la seguridad? Efectivamente lo podría ser si no mantenemos un mínimo de atención, ya que en este tipo de programas se pueden producir desbordamientos de búfer que comprometan su sistema. Recuerde siempre lo siguiente:

No asignar el bit SUID salvo cuando sea estrictamente necesario.
Comprobar que cualquier programa con est bit activo no tiene ningún desbordamiento de buffer (conocido).
No asignarlo jamás si el programa permite salir a la shell.
Atributo SGID: (Para ficheros)
Si está activo en los permisos de grupo, este bit controla el estado de “poner id de grupo” de un fichero. Actúa de la misma forma que SUID, salvo que afecta al grupo. El fichero tiene que ser ejecutable para que esto tenga algún efecto.
Atributo SGID: (Para directorios)
Si activa el bit SGID en un directorio (con “chmod g+s directorio”), los ficheros creados en ese directorio tendrán puesto su grupo como el grupo del directorio.
A continuación se describen los significados de los permisos de acceso individuales para un fichero. Normalmente un fichero tendrá una combinación de los siguientes permisos:

-r——– Permite acceso de lectura al propietario
–w——- Permite modificar o borrar el fichero al propietario
—x——- Permite ejecutar este programa al propietario, (los guiones de shell también requieren permisos de lectura al propietario)
—s——- Se ejecutará con usuario efectivo ID = propietario
———-s– Se ejecutará con usuario efectivo ID = grupo
-rw———T No actualiza “instante de última modificación”. Normalmente usado para ficheros de intercambio (swap)
—t——- No tiene efecto. (antes sticky bit)

A continuación se describen los significados de los permisos de acceso individuales para un directorio. Normalmente un directorio tendrá una combinación de los siguientes permisos:

dr——– Permite listar el contenido pero no se pueden leer los atributos.
d–x——- Permite entrar en el directorio y usar en las rutas de ejecución completas.
dr-x——- Permite leer los atributos del fichero por el propietario.
d-wx——- Permite crear/borra ficheros.
d———x-t Previene el borrado de ficheros por otros con acceso de escritura. Usado en /tmp

d—s–s– No tiene efecto

Los ficheros de configuración del sistema (normalmente en /etc) es habitual que tengan el modo 640 (-rw-r——), y que sean propiedad del root. Dependiendo de los requisitos de seguridad del sistema, esto se puede modificar. Nunca deje un fichero del sistema con permiso de escritura para un grupo o para otros. Algunos ficheros de configuración, incluyendo /etc/shadow, sólo deberían tener permiso de lectura por root, y los directorios de /etc no deberían ser accesibles, al menos por otros.

SUID Shell Scripts
Los scripts de shell SUID son un serio riesgo de seguridad, y por esta razón el núcleo no los acepta. Sin importar lo seguro que piense que es su script de shell, puede ser utilizado para que un cracker pueda obtener acceso a una shell de root.

2. Herramientas de transferencia de archivos

FTP

En informática, de transferencia de archivos es un término genérico para referirse al acto de transmisión de ficheros a través de una red informática. Si bien el término “transferencia de archivos” suele estar ligado al Protocolo de Transferencia de Archivos (FTP)

FTP (siglas en inglés de File Transfer Protocol, ‘Protocolo de Transferencia de Archivos’) en informática, es un protocolo de red para la transferencia de archivos entre sistemas conectados a una red TCP (Transmission Control Protocol), basado en la arquitectura cliente-servidor.

Desde un equipo cliente se puede conectar a un servidor para descargar archivos desde él o para enviarle archivos, independientemente del sistema operativo utilizado en cada equipo.
El servicio FTP es ofrecido por la capa de aplicación del modelo de capas de red TCP/IP al usuario, utilizando normalmente el puerto de red 20 y el 21. Un problema básico de FTP es que está pensado para ofrecer la máxima velocidad en la conexión, pero no la máxima seguridad, ya que todo el intercambio de información, desde el login y password del usuario en el servidor hasta la transferencia de cualquier archivo, se realiza en texto plano sin ningún tipo de cifrado, con lo que un posible atacante puede capturar este tráfico, acceder al servidor y/o apropiarse de los archivos transferidos.
Para solucionr este problema son de gran utilidad aplicaciones como scp y sftp, incluidas en el paquete SSH, que permiten transferir archivos pero cifrando todo el tráfico.

Concretando:
El FTP es un sistema que nos permite de forma cómoda subir o bajar archivos a otra ubicación. Tiene que quedar claro varias cosas.

Hay FTP públicos y privados, es decir en algunos todo el mundo puede entrar y en otros necesitas un nombre de usuario y contraseña.
No todas las opciones te estarán permitadas, hay algunos FTP que solo puedes bajar y en otros te dan la opción de subir en algún directorio.
Debes conocer la reglas de ese FTP.
Debes conocer la dirección del FTP servidor, así como los datos de acceso.
El FTP se basa en la comunicación de dos “ordenadores”, uno es el CLIENTE y otro el SERVIDOR, el cliente con un software se conecta al servidor y se establece una conexión.

Herramientas de transferencia de archivos

Lo normal es desarrollar nuestro sitio de forma local, es decir, en nuestro propio equipo. Pero una vez conseguimos un servidor, gratuito o de pago, tenemos que subir a él todos nuestros archivos.

Hay varias opciones a la hora de subir los archivos al servidor:

Utilizar un programa específico para ello como el Filezilla.

Se trata de un cliente FTP. Para transferir ficheros por FTP se necesita tener instalado un programa cliente de FTP en nuestro ordenador y conocer la dirección del servidor FTP con el que queremos conectar. Nuestra dirección nos la proporcionará la empresa con la que tengamos el alojamiento.
Una vez conectado, los programas FTP tienen un interfaz bastante simple. Básicamente, nos muestran los archivos de nuestro equipo a un lado, y los del servidor a otro. Con lo que no tenemos más que arrastrar y soltar archivos de un lado a otro.
La principal ventaja del FTP, a parte de su comodidad, es que nos permite subir varios archivos y carpetas enteras a la vez, sin tener que ir uno por uno.

Utilizar un editor de páginas web que incorpore esta funcionalidad.

Algunos editores profesionales incorporan el acceso FTP. Esto que nos permite subir y descargar los archivos cómodamente usando el interfaz del editor. Además, nos permiten tenerlos sincronizados, e ir subiendo automáticamente aquellos archivos que vayamos actualizando.

Utilizar las facilidades proporcionadas por la empresa de hosting.

Realmente no necesitamos ninguno de los programas anteriores, aunque casi siempre sean de mejor calidad. Toda empresa de hosting nos proporciona un interfaz a través del que subir archivos al servidor. Este interfaz depende de la empresa. Los de alojamientos gratuitos suelen ser bastantes simples, permitiéndonos subir unos pocos archivos cada vez a través de un formulario web. Esto resulta muy tedioso si tenemos que subir un sitio completo de tamaño medio. Los alojamientos de pago suelen tener un mejor servicio. Algunos implementan un interfaz muy completo (WebFTP) que realmente puede sustituir al cliente FTP.

Acceder por medio del navegador

Utilizando el protocolo ftp

Filezilla:

FileZilla es un cliente FTP multiplataforma de código abierto y software libre, licenciado bajo la Licencia Pública General de GNU. Soporta los protocolos FTP, SFTP y FTP sobre SSL/TLS (FTPS). Inicialmente fue diseñado para funcionar en Microsoft Windows, pero desde la versión 3.0.0, gracias al uso de wxWidgets, es multiplataforma, estando disponible además para otros sistemas operativos, entre ellos GNU/Linux, FreeBSD y Mac OS X.
El código fuente de FileZilla y las descargas estaban2 hospedadas en SourceForge, el cual presentó a FileZilla como el Proyecto del Mes en noviembre de 2003.3 Actualmente hospeda el código fuente en su propio sitio web y las descargas en Ohloh.

De acuerdo con la documentación de ayuda, FileZilla comenzó siendo un proyecto de clase de informática en enero de 2001 de Tim Kosse y dos compañeros de clase. La versión alfa fue lanzada

finales de febrero de 2001, y todas las características requeridas se incorporaron en la beta 2.1.
Características

Administrador de sitios: permite a un usuario crear una lista de sitios FTP con sus datos de conexión, como el número de puerto a usar, o si se utiliza inicio de sesión normal o anónima. Para el inicio normal, se guarda el usuario y, opcionalmente, la contraseña.

Registro de mensajes: se muestra en la parte superior de la ventana. Muestra en forma de consola los comandos enviados por FileZilla y las respuestas del servidor remoto.

Vista de archivo y carpeta: situada en la parte central de la ventana, proporciona una interfaz gráfica para FTP. Los usuarios pueden navegar por las carpetas, ver y alterar sus contenidos tanto en la máquina local como en la remota, utilizando una interfaz de tipo árbol de exploración. Los usuarios pueden arrastrar y soltar archivos entre los ordenadores local y remoto.

Cola de transferencia: situada en la parte inferior de la ventana, muestra en tiempo real el estado de cada transferencia activa o en cola.

2.1. Parámetros de Configuración de la transferencia de archivos

Los comandos FTP

Toda comunicación que se realice en el canal de control sigue las recomendaciones del protocolo Telnet. Por lo tanto, los comandos FTP son cadenas de caracteres Telnet (en código NVT-ASCII) que finalizan con el código de final de línea Telnet (es decir, la secuencia <CR>+<LF>, Retorno de carro seguido del carácter Avance de línea indicado como <CRLF>).
Si el comando FTP tiene un parámetro, éste se separa del comando con un espacio (<SP>).

Los comandos FTP hacen posible especificar:

El puerto utilizado
El método de transferencia de datos
La estructura de datos
La naturaleza de la acción que se va a realizar (Recuperar, Enumerar, Almacenar, etc.)

Existen tres tipos de comandos FTP diferentes:

Comandos de control de acceso
Comandos de parámetros de transferencia
Comandos de servicio FTP

Comandos de control de acceso	
Comando	**Descripción**
USER	Cadena de caracteres que permite identificar al usuario. La identificación del usuario es necesaria para establecer la comunicación a través del canal de datos.
PASS	Cadena de caracteres que especifica la contraseña del usuario. Este comando debe ser inmediatamente precedida por el comando USER. El cliente debe decidir si esconder la visualización de este comando por razones de seguridad.
ACCT	Cadena de caracteres que especifica la cuenta del usuario. El comando generalmente no es necesario. Durante la respuesta que acepta la contraseña, si la respuesta es 230, esta etapa no es necesaria; Si la respuesta es 332, sí lo es.

CWD	Change Working Directory (Cambiar el directorio de trabajo): este comando permite cambiar el directorio actual. Este comando requiere la ruta de acceso al directorio para que se complete como un argumento.
CDUP	Change to Parent Directory (Cambiar al directorio principal): este comando permite regresar al directorio principal. Se introdujo para resolver los problemas de denominación del directorio principal según el sistema (generalmente "...").
SMNT	Structure Mount (Montar estructura):
REIN	Reinitialize (Reinicializar):
QUIT	Comando que permite abandonar la sesión actual. Si es necesario, el servidor espera a que finalice la transferencia en progreso y después proporciona una respuesta antes de cerrar la conexión.

Comandos de parámetros de transferencia	
Comando	**Descripción**
PORT	Cadena de caracteres que permite especificar el número de puerto utilizado.
PASV	Comando que permite indicar al servidor de DTP que permanezca a la espera de una conexión en un puerto específico elegido aleatoriamente entre los puertos disponibles. La respuesta a este comando es la dirección IP del equipo y el puerto.
TYPE	Este comando permite especificar el tipo de formato en el cual se enviarán los datos.
STRU	Carácter Telnet que especifica la estructura de archivos (F de File [Archivo], R de Record [Registro], P de Page [Página]).
MODE	Carácter Telnet que especifica el método de transferencia de datos (S de Stream [Flujo], B de Block [Bloque], C de Compressed [Comprimido]).

Comandos de servicio FTP	
Comando	**Descripción**
RETR	Este comando (RETRIEVE [RECUPERAR]) le pide al servidor de DTP una copia del archivo cuya ruta de acceso se da en los parámetros.
STOR	Este comando (store [almacenar]) le pide al servidor de DTP que acepte los datos enviados por el canal de datos y que los almacene en un archivo que lleve el nombre que se da en los parámetros. Si el archivo no existe, el servidor lo crea; de lo contrario, lo sobrescribe.
STOU	Este comando es idéntico al anterior, sólo le pide al servidor que cree un archivo cuyo nombre sea único. El nombre del archivo se envía en la respuesta.
APPE	Gracias a este comando (append [adjuntar]) los datos enviados se concatenan en el archivo que lleva el nombre dado en el parámetro si ya existe; si no es así, se crea.
ALLO	Este comando (allocate [reservar]) le pide al servidor que reserve un espacio de almacenamiento lo suficientemente grande como para recibir el archivo cuyo nombre se da en el argumento.
REST	Este comando (restart [reiniciar]) permite que se reinicie una transferencia desde donde se detuvo. Para hacer esto, el comando envía en el parámetro el marcador que representa la posición en el archivo donde la transferencia se había interrumpido. Después de este comando se debe enviar inmediatamente un comando de transferencia.

RNFR	Este comando (rename from [renombrar desde]) permite volver a nombrar un archivo. En los parámetros indica el nombre del archivo que se va a renombrar y debe estar inmediatamente seguido por el comando RNTO.
RNTO	Este comando (rename from [renombrar a]) permite volver a nombrar un archivo. En los parámetros indica el nombre del archivo que se va a renombrar y debe estar inmediatamente seguido por el comando RNFR.
ABOR	Este comando (abort [cancelar]) le indica al servidor de DTP que abandone todas las transferencias asociadas con el comando previo. Si no hay conexión de datos abierta, el servidor de DTP no realiza ninguna acción; de lo contrario, cierra la conexión. Sin embargo, el canal de control permanece abierto.
DELE	Este comando (delete [borrar]) permite que se borre un archivo, cuyo nombre se da en los parámetros. Este comando es irreversible y la confirmación sólo puede darse a nivel cliente.
RMD	Este comando (remove directory [eliminar directorio]) permite borrar un directorio. El nombre del directorio que se va a borrar se indica en los parámetros.
MKD	Este comando (make directory [crear directorio]) permite crear un directorio. El nombre del directorio que se va a crear se indica en los parámetros.
PWD	Este comando (print working directory [mostrar el directorio actual]) hace posible volver a enviar la ruta del directorio actual completa.
LIST	Este comando permite que se vuelva a enviar la lista de archivos y directorios presentes en el directorio actual. Esto se envía a través del DTP pasivo. Es posible indicar un nombre de directorio en el parámetro de este comando. El servidor de DTP enviará la lista de archivos del directorio ubicado en el parámetro.
NLST	Este comando (name list [lista de nombres]) permite enviar la lista de archivos y directorios presentes en el directorio actual.
SITE	Este comando (site parameters [parámetros del sistema]) hace que el servidor proporcione servicios específicos no definidos en el protocolo FTP.
SYST	Este comando (system [sistema]) permite el envío de información acerca del servidor remoto.
STAT	Este comando (Estado: [estado]) permite transmitir el estado del servidor; por ejemplo, permite conocer el progreso de una transferencia actual. Este comando acepta una ruta de acceso en el argumento y después devuelve la misma información que LISTA pero a través del canal de control.
HELP	Este comando permite conocer todos los comandos que el servidor comprende. La información se devuelve por el canal de control.
NOOP	Este comando (no operations [no operación]) sólo se utiliza para recibir un comando OK del servidor. Sólo se puede utilizar para no desconectarse después de un período de inactividad prolongado.

Las respuestas FTP

Las respuestas FTP garantizan la sincronización entre el cliente y el servidor FTP. Por lo tanto, por cada comando enviado por el cliente, el servidor eventualmente llevará a cabo una acción y sistemáticamente enviará una respuesta.

Las respuestas están compuestas por un código de 3 dígitos que indica la manera en la que el comando enviado por el cliente ha sido procesado. Sin embargo, debido a que el código de 3 dígitos resulta difícil de leer para las personas, está acompañado de texto (cadena de caracteres Telnet

separada del código numérico por un espacio).

Los códigos de respuesta están compuestos por 3 números, cuyos significados son los siguientes:

El primer número indica el estatuto de la respuesta (exitosa o fallida)
El segundo número indica a qué se refiere la respuesta.
El tercer número brinda un significado más específico (relacionado con cada segundo dígito).

Primer número		
Dígito	**Significado**	**Descripción**
1yz	Respuesta positiva preliminar	La acción solicitada está en progreso. Se debe obtener una segunda respuesta antes de enviar un segundo comando.
2yz	Respuesta de finalización positiva	La acción solicitada se ha completado y puede enviarse un nuevo comando.
3yz	Respuesta intermedia positiva	La acción solicita está temporalmente suspendida. Se espera información adicional del cliente.
4yz	Respuesta de finalización negativa	La acción solicitada no se ha realizado debido a que el comando no se ha aceptado temporalmente. Se le solicita al cliente que intente más tarde.
5yz	Respuesta negativa permanente	La acción solicitada no se ha realizado debido a que el comando no ha sido aceptado. Se le solicita al cliente que formule una solicitud diferente.

Segundo número		
Dígito	**Significado**	**Descripción**
x**0**z	Sintaxis	La acción tiene un error de sintaxis o sino, es un comando que el servidor no comprende.
x**1**z	Información	Ésta es una respuesta que envía información (por ejemplo, una respuesta a un comando STAT).
x**2**z	Conexiones	La respuesta se refiere al canal de datos.
x**3**z	Autenticación y cuentas	La respuesta se refiere al inicio de sesión (USUARIO/CONTRASEÑA) o a la solicitud para cambiar la cuenta (CPT).
x**4**z	No utilizado por el protocolo FTP.	
x**5**z	Sistema de archivos	La respuesta se relaciona con el sistema de archivos remoto.

3. Publicación de páginas web

3.1. Buscadores genéricos

Existe una increíble cantidad de información en Internet (algunos miles de millones de documentos) y la mayoría de esta información se actualiza día a día. Por esta razón, un motor de búsqueda es una herramienta esencial para encontrar lo que se necesita.

Motor de búsqueda

Un motor de búsqueda (también llamado Searchbot) es una herramienta hardware y software que indexa páginas Web para que se puedan buscar a través de palabras claves en un formulario de búsqueda.

¿Cómo funciona un motor de búsqueda?

Los robots o spiders (un tipo de software) recorren la Web indexando su contenido dentro enormes de bases de datos que se pueden consultar.

Como ningún motor de búsqueda puede abarcar todas las páginas en un sólo día (generalmente todo el proceso tarda varias semanas), cada motor adopta su propia estrategia para calcular la frecuencia de actualización de los sitios.
Cómo usar un motor de búsqueda

Cuando el usuario de un motor de búsqueda llena el formulario, elige las palabras a buscar (y a veces aquellas que no se van a buscar) con la ayuda de un operador booleano como puede ser “y”, “o”, y “no” (simbolizados por +, - y otros). La solicitud se envía al motor de búsqueda. El motor busca en sus bases de datos cada una de estas palabras y luego delimita la búsqueda quitando las páginas que no coinciden con el criterio.
A continuación, reenvía una lista de vínculos de las páginas incluyendo el comienzo del texto de la página, texto especificado por el creador de la página con unas etiquetas especiales llamadas metatags, e incluso un extracto de la página que contiene las palabras que se estaban buscando.

Estas respuestas se clasifican por relevancia según el criterio del motor de búsqueda, por ejemplo, el porcentaje de palabras que coinciden con la búsqueda, la densidad de las palabras claves (la cantidad de veces que éstas aparecen en la página), etc.
Motor de metabúsqueda

Un “motor de metabúsqueda” es una herramienta de búsqueda que utiliza los resultados de muchos otros motores de búsqueda.

Búsqueda en Internet

Debido a la cantidad de páginas web existentes, se hace necesario emplear una herramienta para encontrar una página específica que se ajuste a sus parámetros de búsqueda: un motor de búsqueda.

Para utilizar un motor de búsqueda, simplemente escriba palabras clave (términos de búsqueda) en el campo apropiado, presione la tecla Aceptar y espere los resultados. Antes de hacerlo, debe determinar qué clase de palabras debe introducir para tener más posibilidades de encontrar la información que busca.

El motor busca páginas que contengan estos términos y páginas que estén vinculadas a otras que usen hipervínculos que contengan estos términos. Sin embargo, los resultados de una búsqueda pueden ser totalmente diferentes si se utilizan palabras clave separadas por espacios, entre comillas o divididas con un operador específico. Por lo tanto, puede ser necesario refinar su búsqueda utilizando palabras clave adicionales y operadores especiales (la tabla que se encuentra más abajo resume los tipos de operadores y sus efectos).

Si no se obtienen resultados y el motor de búsqueda muestra un mensaje que dice “Su búsqueda

no se ajusta a ningún documento", deberá expandir su búsqueda utilizando términos que se relacionen con el tema que busca o eliminando algunos de los que ha empleado.

Si se produce la situación opuesta (demasiados resultados), deberá acotar su búsqueda agregando restricciones, como por ejemplo pedir que los resultados contengan sólo las palabras solicitadas o que excluyan aquellas no deseadas.

Operadores de búsqueda

Tipo de búsqueda	**Solicitud a introducir**
Nombre propio	Escriba el nombre de la persona
Frase	Escriba la frase entre comillas "buscar frase"
Con **todos** los nombres	Use los Booleanos **Y** o **CERCA** para acotar la búsqueda: +nombre1 +nombre2 +nombre3
Con **al menos uno** de los nombres	Use el Booleano **O** nombre1 + nombre2 + nombre3
Que **NO** contenga una palabra	Use el Booleano **NO** nombre1 + nombre2 + nombre3
Nombre con **múltiples finales posibles**	Use truncamiento: nombre* el motor de búsqueda buscará las páginas que contengan las palabras: nombre, nombrado, nombres, sin nombre, etc.
Mayúsculas	Si escribe una palabra en minúscula, el motor remitirá resultados con mayúsculas y minúsculas. Si escribe las letras en mayúscula, el motor remitirá resultados sólo con mayúsculas.

Un ejemplo de búsqueda

Supongamos que quiere buscar páginas web sobre procuradores generales en el motor de búsqueda Google.

Si introduce:
procurador general
,el motor remitirá todas las páginas que contengan la palabra procurador y todas las páginas que contengan la palabra general (así como algunas que contengan ambas, desde luego, pero estarán perdidas entre todos los otros documentos).

Ahora escriba:
procuradores generales
Esta vez, el motor sólo remitirá un número limitado de páginas que contengan las palabras "procuradores generales" (en plural), ya que ha optado por omitir las instancias en singular (procurador general).

Al escribir:
procurador general*
Ahora tiene las palabras procurador general y procuradores generales, pero hay un problema: la búsqueda excluye todas las páginas donde procurador general esté escrito como:

procurador-general
procuradores general
procurador-generales
procuradores-general
procuradores/general
y demás.

El asunto, entonces, es encontrar la manera de remitir todos los usos posibles. Para hacerlo escriba:
+procurador* +general*
Obtendrá todas las páginas que contengan al menos una palabra que empiece con ambas raíces procurador y general.

CONCLUSIÓN: Una búsqueda óptima se debe llevar a cabo para satisfacer sus expectativas; generalmente es diferente de una búsqueda básica, pero ahorra tiempo.

Resumen

Operador Booleano	**Resultado**
+nombre1 +nombre2	*Remite documentos que contengan ambos términos buscados*
nombre1 + nombre2 *o* **nombre1 nombre2**	*Remite documentos con uno (o ambos) términos buscados*
+nombre1 -nombre2	*Remite documentos que no contengan la palabra que sigue al operador -*
nombre*	*Remite documentos que contengan palabras similares al término buscado*
comillas	*Remiten documentos que contengan la frase completa*
Esencial (+)	*El símbolo "+" indica que se debe encontrar una palabra*
Exclusiones (-)	*El símbolo "-" excluye una palabra de la búsqueda*

3.2. Optimización en los motores de búsqueda: SEO (SEARCH ENGINE OPTIMIZATION)

Criterios para un buen indexado en los buscadores

1. Compruebe que su Site es potencialmente indexable por Google.

Cómo puedo comprobarlo:
Para comprobar si la página puede ser indexada, debe utilizar aplicaciones que visualizan su página tal y como lo haría Google. Una de estas aplicaciones es el navegador textual Lynx, también dispone de una versión on-line. Tenga en cuenta que Google sólo podrá acceder al contenido que

Lynx le muestre al analizar su web.

Por qué no podría indexar Google mi página web:

Los principales motivos por los que una página web no puede ser indexada son:

La página está desarrollada exclusivamente en flash: Aunque Google es capaz de leer e indexar páginas flash a partir de la versión 8, lo cierto es que la incapcidad de flash a la hora de jerarquizar y marcar semánticamente el contenido sigue siendo un lastre a la hora de alcanzar un posicionamiento eficiente.

La página contiene frames: Al utilizar frames en el diseño de una web se pierde el concepto de página como unidad de presentación y Google no será capaz de acceder al contenido de nuestra página.

Existen redireccionamientos en javascript: Google sólo recorre los enlaces HTML de una página. Puede verse en la necesidad de hacer una redirección inicial para dirigir al visitante a una u otra página en función de ciertos parámetros como procedencia, idioma,etc. Una simple redirección en javascript en la home puede estar impidiendo a Google acceder. Una solución alternativa válida es realizar la redirección en el servidor.

Utiliza funcionalidades en DHTML: el HTML dinámico (menús desplegables, efectos visuales, etc) no es siempre accesible por los buscadores, por tanto debe asegurarse de que el contenido principal de sus páginas sea siempre accesible.

Trabaja con un gestor de contenido (cms): Si su empresa utiliza un gestor de contenido para gestionar la información de su web, debe asegurarse de que su sistema puede exportar el contenido de manera que los motores de búsqueda puedan indexarlo.

2. Compruebe que Google es capaz de recorrer todas las páginas que conforman el Site.

Cómo puedo comprobarlo:

Podemos recurrir a herramientas especializadas para comprobar la estructura de enlaces de nuestra web. Yo le recomiendo la aplicación gratuita Xenu's Link Sleuth que nos permite realizar un exhaustivo estudio de la estructura interna del Site.

Por qué no podría Google recorrer mi página web:

Cuando Google visita nuestro site, comienza a recorrer todos los enlaces que encuentra, aquellas secciones a las que no pueda acceder no serán indexadas. Existen diversos motivos por los que Google no puede recorrer nuestro Site:

No se puede acceder a todas las secciones desde la home. Tenga en cuenta que la home es el punto de partida hacia el resto de páginas.

Existen links rotos.

Se utiliza javascript para enlazar a alguna sección o página.

Cómo puedo solucionarlo:

Una manera sencilla de asegurarnos de que todas las secciones son accesibles es crear un mapa web. El mapa web de un sitio es una lista jerarquizada de vínculos a todas las páginas del Site.

3. Compruebe que su Site no incumple el código para Webmasters de Google.

Si su Site ha sido desarrollado u optimizado por una empresa SEO debería tener especial atención en comprobar que no han utilizado técnicas de posicionamiento fraudulentas.

Evite los textos y enlaces ocultos.
No cree páginas con contenido irrelevante.
No cree páginas ni subdominios con contenidos duplicados.
No intente hacer cloaking: ofrecer a los buscadores diferente contenido que a sus visitantes con la intención de mejorar su ranking.
Cuide la calidad de sus enlaces, evite aquellos que provengan de webs especializadas en la venta de links y de spamers.
No utilice programas para realizar altas masivas en buscadores.
Evite realizar Sneaky Redirection: conjunto de entre 10 y 20 páginas sin contenido relevante que enlazan entre si. Al acceder a alguna de ellas redirecciona al usuario a otra página distinta.
Evite crear doorways: páginas creadas exclusivamente para los motores de búsqueda
En resumen, evite los trucos mágicos para mejorar el posicionamiento web de su página en Google.

Tenga presente que si nuestra web es penalizada por Google, recuperar la indexación será una tarea muy complicada y en acasiones imposible.

4. Consiga varios enlaces externos de calidad.

Conseguir enlaces de calidad hacia nuestra página web es sin duda la manera más rápida para ser indexados en Google:

Dé de alta su site en los principales directorios:

Algunos directorios permiten el alta manual y gratuita por lo que son un buen lugar donde conseguir un enlace de calidad. Es muy importante que dé de alta su Site en los siguientes directorios:

Open Directory Project (Google directory)
Yahoo!
vlib

Aunque se tiende a pensar lo contrario, los directorios y los buscadores son completamente distintos. Sus enfoques son completamente diferentes ya que registran y presentan la información de forma distinta. Una de las diferencias principales es que los directorios están estructurados temáticamente y requieren que demos de alta nuesta web en la categoría relacionada con nuestros servicios y productos.

Consiga enlaces en sites que ya estén indexados.

Si es el webmaster de una página ya indexada, puede incluir en la home un enlace hacia su nueva web. Hágalo de manera clara, aprovechando la sección de enlaces o incluyendo un banner. Nunca recurra a los pop-ups pues no son accesibles por Google.

Realice campañas de marketing y PR para informar a todos los Sites que podrían estar interesados en que su página está on-line: centros de prensa, portales especializados, páginas del sector, etc.

Tenga en cuenta que un sólo enlace de calidad puede ser suficiente para conseguir la indexación.

5. Crear y dar de alta un Site Map para Google.

Qué es un Google Site Map:

Es un sistema gratuito que nos ofrece Google para que le indiquemos cuáles son las páginas que conforman mi sitio web. De esta manera nos aseguraremos de que Google puede encontrar todos nuestros contenidos de manera sencilla. Además, también nos proporciona estadísticas y resúmenes de posibles errores de indexación.

Cómo se crea un Google Site Map:

Un Google Site map es un fichero que funciona como índice de todas nuestras páginas. Puede tener diversos formatos pero el más habitual es el basado en el estándar XML. Existen muchas herramientas para generar un site map de manera automática. La aplicación xml sitemaps es un generador online y gratuito para webs con menos de 500 páginas.

El proceso para dar de alta el Site Map es sencillo y dispone de mucha documentación en Internet.

Si ha seguido correctamente estos 5 pasos, ahora sólo le queda esperar a que Google encuentre su web e indexe sus páginas. Tenga en cuenta que Google funciona como una araña (si tiene fobia a estos arácnidos piense en otra cosa) que recorre la inmensa tela de webs que conforman Internet.

El tiempo de indexación depende de muchos factores, pero no debería dilatarse más de 15 días. Usted sabrá que la indexación se ha realizado con éxito si al escribir en el campo de búsqueda de Google: site:www.nombre_su_domino_com, el buscador le devuelve todas las páginas que conforman su site.

Buenas prácticas para conseguir mejores posiciones en los resultados de las búsquedas en Wordpress

- Como quieres que se vea y accedan a tu **URL**. Ajustes/Enlaces permanentes (permalinks). Decide la estructura que prefieras para tu URL. Hay una serie de etiquetas disponibles
- **Canonicalización** de tus URLs. Para un buscador no es lo mismo “http://miweb.com” que “http://www.miweb.com”. Si no se define, se considera contenido duplicado, y se penaliza. Ajustes/Generales/ Dirección de WordPress (URL) y Dirección del sitio (URL)
- Incluir **palabras clave, o keywords** en títulos, **metaetiquetas**, en el primer y último párrafo, así como intercaladas a lo largo de la página con distintos sinónimos, hará que esa página sea relevante para los buscadores para esa palabra clave. En este punto hay que tener especial cuidado en no sobreoptimizar pues Google puede penalizarte. Lo importante es que los usuarios lean un contenido de interés bien redactado y a la vez tengamos en cuenta a los buscadores.
- **Imágenes** con palabras clave, tanto en el título de las mismas, como en las descripciones concretas que de cada una de ellas se haga, rellenado con una frase descriptiva que contenga nuestras palabras clave. Asimismo, Todas las imágenes deberían optimarse antes de subirlas para reducir peso, así como que sus dimensiones no excedan en lo posible el tamaño con el que se visualizan.
- **SEO** interno de la página. El mismo se debe estructurar poniendo un título específico en cada página, anteponiendo el nombre del producto por encima del nombre comercial, y utilizando

una descripción somera de hasta 70 caracteres. Igualmente se podrá poner una etiqueta de descripción, de hasta 140 caracteres, en cada página para favorecer el posicionamiento web.

- Utilizar **contenido original**. Los generadores de contenidos, aunque cada vez son más sofisticados, siguen teniendo peores resultados que el contenido humano. De igual forma, hacer una buena inversión en redactores de calidad, optimiazando igualmente la calidad de los textos.
- Colocar imágenes y videos en nuestras páginas hará que el tiempo de permanencia del visitante en él aumente y por tanto Google lo tendrá en cuenta.
- **Mejor páginas que entradas** para el SEO: Los comentarios modifican el valor que tienen las palabras clave que cuidadosamente hemos creado.
- Hay que estar atentos a los enlaces salientes que colocamos en nuestro sitio, es importante que no sean enlaces rotos o que apunten a sitios caidos. Un plugin que puede ayudarte en esto es **Broken Link Checker.**
- **Linkbuilding** o construcción de enlaces, es una de las estrategias del SEO que consiste en conseguir que otras páginas web enlacen a la página que interesa que los buscadores consideren relevante y la posicionen mejor en sus rankings. *http://es.wikipedia.org/wiki/Link_building*
- **Link baiting** (traducido al castellano sería "cebo de enlaces") es un término en inglés que hace referencia a cualquier contenido o característica de un sitio web de que el usuario estimula a los visitantes a crear enlaces hacia él desde sus propias webs. *http://es.wikipedia.org/wiki/Link_bait*
- **Anchor text**: El texto de un texto enlazado. El anchor text es la parte visible de un link. El anchor es de gran utilidad para el posicionamiento ya que es detectado por los buscadores como unos de los factores claves a la hora de determinar la temática de nuestra web, ya que al fin y al cabo, las palabras del anchor son las palabras que otras webs utilizan para definir nuestra web. *http://www.paranovatos.com/anchor-text-wordpress.html*
- **Enlaces** a entradas relacionadas: Se puede instalar un plugin que inserte enlaces en un post de otros más antiguos que tengan relación con este. Buscar plugin por "link related post"
- Es aconsejable crear un fichero **favicon.ico** y situarlo en la carpeta raiz de tu sitio web. Por ejemplo con el plugin RealFaviconGenerator Más info: http://codex.wordpress.org/Creating_a_Favicon#Installing_a_Favicon_in_WordPress
- **Rendimiento** de la página, cuanto más tiempo espera el usuario hasta que la página está cargada, más bajan tus posiciones en las páginas de resultados. Se puede instalar un plugin, W3 Total Cache o **WP Super Cache**. Encontraréis una guía de como usarlo aquí: 12 pasos para mejorar el SEO en WordPress y la experiencia de usuario (paso 6: Optimizar la velocidad de carga de la web)
- Diseño adaptado a dispositivos móviles o **responsive design**: Google recomienda que los sitios web estén adaptados a dispositivos móviles para mejorar la experiencia de navegación de los usuarios. Y te premia aumentando tu posicionamiento en los resultados de búsqueda. Asegúrate de que tu sitio web cuente con un diseño web responsive.
- Revisar y optimizar el **código** de tu tema o plantilla: Una vez que has seleccionado tu tema, sería conveniente que analizaras y revisaras el código del tema para la detección de posibles errores de optimización. Puedes usar alguna extensión de Firefox, por ejemplo **SEO Doctor.**
- Difundir tus contenidos a través de Internet, de tal forma que ayude a obtener tráfico adicional. Existen sitios como **Technorati** y **Sphere** donde la gente navega para encontrar fácilmente las entradas que has publicado recientemente. También deberías de descargar e instalar algún plugin que permita a tus visitantes compartir tus entradas en diversos medios sociales como Facebook, Twitter, LinkeIn, Printerest, etc.
- Estudio de **marketing online** en relación al target de los clientes potenciales que se desean en la web, para articular a partir del mismo todos los puntos anteriores

SEO en Wordpress con "Yoast WordPress SEO"

Eliminar las "stop words" com "y", "a", "en", etc
Optimizar los títulos SEO
https://yoast.com/articles/wordpress-seo/#titles

Envío de mapas XML de manera automática a los principales buscadores
Uso de las "migas de pan" (breadcrumbs): SEO/Enlaces internos
https://yoast.com/articles/wordpress-seo/#bread-crumbs
Texto completo (inglés): The Definitive Guide To Higher Rankings For WordPress Sites

Bibliografia:

https://yoast.com/articles/wordpress-seo/
http://www.ipixelestudio.com/blog/pasos-seo-wordpress-posicionamiento-web.html
http://marketingonlinezaragoza.org/10-buenas-practicas-seo-para-cualquier-sitio-web/
http://www.pauklein.com/que-es-el-anchor-text/
http://www.wordtracker.com/academy/content/wordpress/seo-wordpress-mistakes

3.3. Aplicaciones de publicación automatizada

Aplicaciones gratuitas e incorporadas a servidores gratuitos y de pago.
CMS: Sistema de gestión de contenido (Content Management System)

Un sistema de gestión de contenido(CMS) es una página web con algunas funciones de publicación. En concreto, tiene una interfaz administrativa que permite al administrador del sitio crear u organizar distintos documentos.

En teoría, el CMS debe tener un sistema de flujo de trabajo que permita a un equipo editorial trabajar de manera simultánea y a un director de publicación aprobar las contribuciones antes de que se publiquen en línea.

En principio, los artículos y el contenido del sitio se guardan en una base de datos, en tanto que las plantillas definen el diseño del contenido.

Un CMS estándar muestra un diseño basado en cajas que se organiza, por lo general, en tres columnas. Muchos CMS tienen un canal RSS que se actualiza automáticamente cuando se publican artículos nuevos.
Más información

Las herramientas CMS más utilizadas son:

Wordpress
Enfocado a la creación de cualquier tipo de sitio, aunque ha alcanzado una gran relevancia usado para la creación de blogs (páginas web con una estructura cronológica que se actualiza regularmente).

https://es.wordpress.org/

Joomla
Permite desarrollar sitios web dinámicos e interactivos. Permite crear, modificar o eliminar contenido de un sitio web de manera sencilla a través de un "panel de administración".

https://www.joomla.org/3/es/

Drupal
Es un marco de gestión de contenidos o CMS libre, modular multipropósito y muy configurable que permite publicar artículos, imágenes, archivos y otras cosas u otros archivos y servicios añadidos como foros, encuestas, votaciones, blogs y administración de usuarios y permisos.

http://drupal.org.es/

Magento
Gestor de contenidos web opensource para **comercio electrónico**. Es una solución flexible y escalable con la que se pueden desarrollar prácticamente todo tipo de proyectos e-commerce.

https://magento.com/

Blogger
Es un servicio creado por Pyra Labs, y adquirido por Google en el año 2003, que permite crear y publicar una bitácora en línea. Para publicar contenidos, el usuario no tiene que escribir ningún código o instalar programas de servidor o de scripting.

https://www.blogger.com/

Typo3
Permite realizar enteramente un sitio web de contenidos, con todo lo que eso implica, pero es también un portal. Administra, en particular, la personalización de las páginas según la identidad de los usuarios, es decir sabe integrar una selección de contenidos en una misma página, según los derechos del usuario identificado.

https://typo3.org/