
Redes informáticas e investigación científica

P.M. Lledó Penalva

CEIFE. Universidad Complutense. Madrid.

Introducción

El objeto de este artículo no es otro que presentar un nuevo conjunto de herramientas que pueden ser muy útiles para obtener y procesar información muy variada y que pueden aportar a la investigación un nuevo entorno para desarrollar nuevos métodos de trabajo más cómodos, rápidos, flexibles y potentes.

Con la caída de precios, la disponibilidad y el incremento de la capacidad y prestaciones de los ordenadores desde mediados de los años ochenta, éstos se han convertido en una herramienta muy importante para muchos sectores de trabajo y prácticamente imprescindibles en el área de la investigación. A veces creemos casi imposible haber estado realizando ciertas tareas sin la ayuda de un ordenador. Como en el caso que nos ocupa, la realización de estudios farmacológicos sería una tarea muy dura sin la ayuda de los ordenadores.

Pongamos como ejemplo los trabajos sobre el proyecto *Genoma Humano*. La ingente cantidad de información a procesar sería imposible de abarcar sin la colaboración de las máquinas. No sólo de ordenadores sino también de analizadores, secuenciadores, etc., que se conectan a los ordenadores. Los investigadores usan los ordenadores no sólo para procesar los datos, sino también para intercambiar la nueva información con otros grupos de investigadores. Por cada segmento de la secuencia de ADN hay un grupo de investigadores trabajando. Todos los grupos están coordinados por un equipo central, que a su vez mantiene una base de datos común con la secuencia de ADN (mapa del genoma humano) descubierta hasta el momento. Este mapa ya está siendo consultado a través de la red por cientos de investigadores que desarrollan terapias basadas en modelos genéticos.

Redes

Ya sabemos que los ordenadores son muy útiles si se usan correctamente, pero ¿por qué son necesarias las «redes de ordenadores»? Como ya hemos citado en alguna ocasión, un orde-

nador aislado, *autista*, es un ordenador limitado a sus propios recursos. Las redes más sencillas son las redes departamentales (LAN: *Local Area Network*), que agrupan ordenadores que realizan unas tareas similares usando casi siempre los mismos recursos (impresoras, aplicaciones, ficheros, etc.) accediendo a ordenadores (servidores) que están dedicados a centralizar los datos y servicios comunes. Estas redes sencillas suelen conectar los ordenadores entre sí por medio de cables de varias clases, que pueden alcanzar longitudes de hasta 200 o 400 m según el tipo de red. Otras redes más amplias, que sirven para conectar departamentos entre sí u otros grupos de máquinas más amplios, se denominan WAN: *Wide Area Network*. Los medios físicos para la conexión suelen ser fibra óptica, módems y cable telefónico, satélites, enlaces de microondas, etc.

Un aspecto a tener en cuenta a la hora de trabajar con redes es la protección de la información. Si queremos que nuestra información esté disponible para los demás, o por lo menos cierta parte de la información, debemos protegerla de accesos indeseados. Para ello cada usuario de un sistema multiusuario tiene una «cuenta» con un nombre de acceso y una palabra clave o *password*. Existen otras formas de proteger la información que se salen del objetivo de esta publicación, como la encriptación y la autenticación.

Como tema central de este artículo vamos a intentar mostrar cómo los recursos de las redes globales son una herramienta fundamental y muy útil para los grupos de investigación. Como algunos ejemplos de redes de ordenadores a nivel mundial contamos con: *Comuserve* (comercial), *FidoNet* (amateur) e *Internet* (semipública).

Internet

Internet es *la red*. Nos atreveríamos a afirmar que es la mayor estructura informática interconectada. Abarca todo el planeta, extendiéndose por unos 130 países de los 5 continentes. En ella se conectan ordenadores de muy variadas características y funcionalidades. Algunos de estos ordenadores se dedican exclusivamen-

te a conectar unos con los otros, a intercambiar correo electrónico u otro tipo de información y datos entre diversas redes. Por eso se conoce a Internet, como la «red de redes». Actualmente también se la conoce como «las autopistas de la información», aunque en algunos países y/o zonas se circule a mayor velocidad que en otros. Otros ordenadores se encargan de recopilar en índices especializados la información que existe en otros ordenadores. Para ello, se usan programas que recorren las redes buscando información significativa en los ficheros de los bancos de datos públicos y otras fuentes de datos, recopilan dicha información y generan índices que serán utilizados a su vez por otros programas buscadores (véase posteriormente).

Internet puede definirse como la combinación de cuatro elementos:

1. Una «red de redes» de ordenadores conectados entre sí que utiliza los protocolos de comunicación TCP/IP para transportar información.

2. Un conjunto de recursos a los que se puede acceder desde esas redes.

3. Un conjunto de servicios que permiten acceder a los recursos disponibles en la red.

4. Una comunidad de gente que utiliza esos recursos y desarrolla dichas redes.

Internet tuvo su origen en una red de ordenadores militares de los EE.UU. Su nombre primitivo fue *ARPANet*. Sobre esta red se fueron desarrollando y conectando otras redes, al principio de otros organismos estatales y no comerciales, y más tarde con organizaciones comerciales americanas y del resto del mundo.

A Internet se conectan otras grandes redes mundiales, tanto comerciales como públicas o amateurs. Como ejemplo, sirvan las anteriormente nombradas *Compuserve* y *FidoNet*, aunque hoy día cualquier otra red tiene una conexión con Internet para facilitar a sus miembros el acceso a los recursos de «la red».

Internet se apoya en una estructura de ordenadores basados en sistemas abiertos, es decir, que su *hardware* puede ser de cualquier marca y el *software* es muy variado. Sólo tienen que ajustarse a ciertas normas y estándares para poder hacer compatibles las aplicaciones, los protocolos de comunicación y los programas de control. Hoy día, el coste de crear un nuevo nodo en Internet es mínimo, haciendo posible que cada departamento o unidad de investigación tenga y maneje el suyo.

El mantenimiento y financiación de Internet es complejo y es un aspecto que escapa a los

objetivos de este artículo. Sí podemos comentar que los recursos públicos son de libre acceso, costando únicamente la conexión a la red. Una vez conectado, «viajar» y coger información es, por lo general, sin costes.

Herramientas

Es obvio que para manejarse entre toda la abundancia de recursos y fuentes de información haya que usar un conjunto de herramientas que nos permita movernos, buscar y seleccionar las vías de acceso y los datos con los que queremos trabajar, así como qué aplicaciones hemos de usar para cada propósito. Las herramientas que comentamos a continuación son universales o también llamadas multiplataforma, es decir, suelen funcionar en varias clases de ordenadores que se conectan a Internet. Pueden funcionar sobre muchas versiones de distintos sistemas operativos tales como unix, Macintosh, MS-DOS, OS/2, VMS, Windows, etc. He aquí las más importantes.

Telnet

Si queremos acceder desde nuestro ordenador habitual a otro ordenador en el que queramos trabajar usamos *Telnet*. De este modo utilizamos la pantalla de nuestro ordenador como si fuese la pantalla del ordenador al que acabamos de conectarnos. A esto se le conoce como «abrir una sesión remota». El ordenador local abre un canal de comunicación con el ordenador remoto y nos permite hacer uso del teclado y pantalla local como si fuese la del remoto. Con ello podemos acceder, por ejemplo, al ordenador central de la biblioteca de la universidad y consultar bibliografía.

Email

Se utiliza para intercambiar mensajes y datos con un usuario predeterminado. Conociendo la dirección de correo electrónico de alguien podemos enviarle un mensaje que puede ser un texto o incluso contener datos binarios codificados; podríamos entonces enviar documentos de aplicaciones como hojas de cálculo, bases de datos, procesadores de texto, etc. Este mensaje puede ir saltando de ordenador en ordenador hasta llegar al ordenador del usuario destinatario, tal y como lo haría una carta de oficina en oficina de correos. Los mensajes pueden estar protegidos contra lectura indeseada por una palabra clave.

News

Es parecido a *email*, pero aquí los mensajes que escribimos serán leídos por todos los suscritos al grupo de *news* al que hemos enviado el mensaje. Igualmente, si estamos interesados en leer noticias sobre un tema en particular deberemos suscribirnos al grupo que más se asemeje a lo que vamos buscando. Por ejemplo: si deseamos leer *news* sobre el proyecto genoma humano nos apuntaremos al grupo:

bionet.molbio.genome-program

FTP

Se usa para acceder a servidores de ficheros remotos. Se establece una sesión de comunicación entre 2 ordenadores, permitiendo al usuario consultar el contenido de los directorios *públicos* del servidor. Se puede solicitar la transmisión de ficheros a nuestro ordenador local. Se puede acceder a los directorios privados del servidor a través de una cuenta de usuario con *password*.

En el caso de nuestro departamento (CEIFE) recibimos los nuevos datos desde Italia vía *FTP*. De este modo, actualizamos la base de datos con un coste mínimo y una gran rapidez. Además, siendo grandes cantidades (40 Mbytes) de información actualizada, la recibimos casi instantáneamente y en modo desatendido.

Archie

Está basado en *FTP*. Es un sistema dedicado a recopilar el contenido de directorios públicos de otros ordenadores. De este modo, cuando necesitamos un fichero en particular, conectamos con un sistema *Archie*, le interrogamos por nombre de fichero y nos devuelve una lista de servidores de ficheros, nombres de directorio y nombres de ficheros por los que buscar.

Gopher

Es un sistema de recopilación de información documental, al que se accede mediante un sistema ramificado de menús. Cada menú te guía a un documento o a otro menú, que puede estar en la misma máquina o en otra.

WWW

World Wide Web es el sistema de información más moderno y sofisticado que usa Internet para búsquedas de información y oferta de ser-

vicios. Hablaremos de él posteriormente en profundidad.

WAIS

Es un sistema, podríamos decir, intermedio entre *Archie* y *Gopher* pero basado en páginas *Web*. Se encarga de explorar las páginas creando índices de todas las palabras significativas permitiéndonos más tarde encontrar páginas que contengan dichas palabras.

Robots, Worms, Spiders y Crawlers

Son sistemas que se encargan de recorrer la Internet saltando de recurso en recurso, guardando las direcciones y las palabras clave de cada recurso de información. Van generando macroíndices, que después serán explorados a su vez por otros programas para buscar información que corresponda a una pregunta del tipo: quiero saber dónde puedo encontrar información acerca de investigación médica sobre fármacos. Para ello la pregunta sería: *health + research + pharma* (depende del sistema que usemos para búsqueda). El sistema nos contesta con una lista de recursos, con un pequeño resumen de la página, en donde podemos consultar al respecto. Estos sistemas son unas de las herramientas más útiles y potentes de la Internet.

Aplicaciones front-end

Son, por ejemplo, los programas que se usan para acceder a recursos de información remotos, tipo bases de datos documentales, como Medline (bibliografía). Se usa un programa cerrado para la comunicación por protocolos TCP/IP con el ordenador remoto que contiene la información.

Compartir recursos

Toda esta macroestructura de vías de información, herramientas, bancos de datos y aplicaciones de búsqueda y manejo de información tiene como objetivos fundamentales proveer de una fuente de información pública, fiable, actualizada y fácilmente consultable a cualquier investigador que tenga acceso a Internet. La potencia y flexibilidad de estas herramientas hace posible actualmente lo que hace unos años era ciencia ficción. Como veremos a continuación, las búsquedas de información por Internet son sencillas y potentes. Encontramos información

que «se dice obsoleta» de sólo 2 meses de antigüedad.

Los usuarios de Internet ponen su información accesible al resto del mundo para que todos puedan compartir sus conocimientos, ver los resultados de sus investigaciones y hacer comentarios al respecto, refutando o aceptando los hechos. Cuando se pretende comenzar un trabajo, lo primero que se debe hacer es comprobar si ya hay alguien trabajando en lo mismo o ya lo ha hecho, para no reinventar la rueda. Se puede entonces seguir la línea de investigación o desviarse por nuevos cauces. El trabajo de investigación en la «aldea global» nos lleva a formar «grupos de investigación virtuales» en los que sus miembros no tienen por qué conocerse personalmente, pero pueden colaborar en proyectos cuya vía de enlace es Internet. Enviando protocolos de investigación vía *email*, recibiendo ficheros de datos por *FTP*, leyendo los últimos comentarios en los grupos de *news*, publicando los resultados de sus trabajos en páginas *Web*, etc. disponemos de una plataforma de trabajo tan potente y dinámica como nunca habíamos tenido oportunidad.

WWW

WWW es el acrónimo de *World Wide Web*, que podría traducirse como telaraña a nivel mundial. Hace referencia a la estructura de Internet, en la que los ordenadores se conectan unos a otros por múltiples vías.

Usa toda la infraestructura de Internet, pero se apoya en la tecnología del *Hypertext Markup Language* (HTML), un lenguaje de programación de páginas de información con formato hipertexto, que permite visualizar información de tipo texto y gráficos. Incluso reproduce sonidos y vídeo en máquinas con *hardware* que lo permitan. En dichas páginas la información puede leerse como en una página de un procesador de textos normal, pero en algunas zonas encontramos **enlaces** (en inglés, *hyperlinks*) con otras partes del texto. Los enlaces, que son palabras o frases en negrita y subrayadas, nos amplían la información a la que estábamos accediendo y nos llevan a nueva información que puede estar localizada en la misma página, en otra página del mismo ordenador o en otra página de otro ordenador remoto. Este es el secreto del WWW. La información está distribuida y relacionada por proximidad temática. Vamos buscando y obteniendo la información necesaria «a demanda».

Sobre las páginas *Web* se construyen a su vez nuevas aplicaciones, ya que el lenguaje HTML

es programable. Podemos detectar desde dónde se nos está pidiendo información, facilitándola en el formato e idioma adecuado en cada momento. Podemos pedir información al «visitante» para actualizar una base de datos o bien si el visitante nos solicita información de cualquier tipo podemos pedirle la dirección de correo o correo electrónico para enviarla haciéndole rellenar un cuestionario en la propia página *Web* que está viendo en su ordenador. También podemos presentar una lista de nuestros servicios o productos, o en nuestro caso una lista de los *abstracts* de los trabajos publicados.

Desde las páginas *Web* podemos lanzar programas que interroguen una base de datos y devolver el resultado de la búsqueda como otra página en el mismo momento. Creamos una nueva página de forma instantánea automáticamente. Las posibilidades del HTML son enormes, y como ejemplos podemos citar: búsquedas de contenidos de otras páginas *Web* de todo el mundo (*Yahoo*, *Lycos*, etc.); tutoriales de cualquier tipo; búsquedas de reacciones adversas en bases de datos farmacológicas; presentaciones de productos comerciales; búsquedas de audición de maquetas de música; visualización de secuencias de vídeo; búsquedas de alojamientos y restaurantes, etc.

Futuro

Podemos jugar a prever el porvenir de los métodos de investigación que se servirán de la red Internet. Ya hemos comentado algo acerca de los *grupos virtuales de investigación* en los que la comunicación a cualquier nivel es por Internet. Podríamos aventurar que es una forma de *teletrabajo*. Esta relación laboral puede motivar que, de vez en cuando, sea necesario realizar una sesión conjunta cara a cara. Podemos organizar una *teleconferencia* a través de Internet.

De igual modo que la aparición de los diarios virtuales en Internet, que están haciendo que las empresas que editan los diarios en papel ya estén editando páginas sobre Internet, las nuevas vías de publicación de los resultados de trabajos de investigación podrían ser las de *autopublicar* en los servidores de información departamental. De este modo, los resultados pueden ser inmediatamente accesibles a todo el mundo, en un formato, además, procesable y editable para que sea más cómodo. Estos *trabajos autopublicados* podrían haber estado revisados por expertos en la materia que habrían recibido el original por *email* y contestado al autor en un muy breve plazo de tiempo. La pu-

blicación después de la revisión es casi inmediata, ahorrando así mucho del tiempo que transcurre entre el envío para publicación y la disponibilidad en papel. La calidad de estos trabajos vendría refrendada por dichos revisores y sería la comunidad científica quien juzgaría al final la calidad y la relevancia del trabajo.

Direcciones útiles

Organizaciones de Salud

CEIFE-Universidad Complutense de Madrid
<http://www.ceife.ucm.es>
 NIH. Instituto Nacional de la Salud de EE.UU.
<http://www.nih.gov/>
 World Health Organization. WHO. OMS.
<http://www.who.org/>
 PAHO-Pan-American Health Organization
<http://www.paho.org/>
 FDA-Food and Drug Administration
<http://www.fda.gov/fdahomepage.html>
 GalaxyNet: Health
<http://www.einet.net/galaxy/Community/Health.html>
 PharmInfo
<http://pharminfo.com/>
 American Medical Association
<http://www.ama-assn.org/>
 Virtual Hospital
<http://vh.radiology.uiowa.edu/>
 PharmWeb
<http://www.mcc.ac.uk/pharmweb/>
 New England Epidemiology Institute
<http://epidemiology.com/>
 ISPE-International Society for
 Pharmacoepidemiology
<http://www.kumc.edu/ISPE/>

Revistas

British Medical Journal
<http://www.bmj.com/bmj/>

Otras direcciones

Yahoo. Un buscador y catalogador de información en páginas Web. Junto con Lycos, InfoSeek y WebCrawler son los más conocidos.
<http://www.yahoo.com/Health/>

Glosario

ARPANET: *Advanced Research Projects Agency Network*. Red de la agencia de proyectos de investigación avanzados de EE.UU.

Browser: programa cliente del *World Wide Web*. Permite «viajar» por Internet explorando sus recursos y accediendo de una manera fácil, dinámica e interactiva a los servicios de la red. Usa de las capacidades hipertexto (HTML) y de los protocolos *HTTP*.

Cliente: un programa que solicita servicios de otro programa (servidor) que los sirve.

Email: es el acrónimo de *electronic mail* (correo electrónico). La gran mayoría de los usuarios de máquinas conectadas a Internet hacen uso del correo electrónico para intercambiar mensajes con otros usuarios en todo el mundo. Los sistemas de correo se encargan de encaminar los mensajes de máquina en máquina hasta llegar a la máquina destino, como lo haría un sistema de correo de oficina en oficina.

Email address: significa dirección de correo electrónico (véase *email*).

FTP: *File Transfer Protocol*. Es una norma o protocolo para copiar ficheros de un ordenador a otro.

HTML: *HyperText Markup Language*. Lengua-je hipertexto basado en marcas. Es el lenguaje con el que se escriben las páginas Web.

HTTP: *HyperText Transport Protocol*. Protocolo de transporte de información hipertexto. Es la base del *World Wide Web*.

Internet: la red mundial de ordenadores conectados mediante protocolos *TCP/IP*.

ListServ: servidor de listas de correo electrónico. Se encarga de contestar automáticamente mensajes de los suscriptores de una (o varias) lista de correo.

Mailing list: lista de correo. Conjunto de direcciones de correo que tienen en común su participación en algún tema de diálogo en particular. Permite poner en contacto a personas o instituciones que participan en un foro de debate común. Para participar se usa un programa para leer correo y para contestar hay que enviar un mensaje al *ListServ*.

Malla Multimedia Mundial (véase *World Wide Web*).

Newsgroup: grupo de *news*. Es similar a la mailing list, pero a un nivel más amplio. Forma parte de *Usenet*. Básicamente se diferencian en el número de gente que participa y en las aplicaciones informáticas que lo hacen posible. Las *newsgroup* son más activas y más globales. El acceso es más sencillo. Para leerlas se usan programas para leer *News*.

Server: servidor de información o de ficheros. Un ordenador que está conectado a Internet para facilitar recursos a clientes de toda la red.

TCP/IP: *Transmission Control Protocol/Internet Protocol*. Es básicamente un conjunto de protocolos de control y transmisión de información entre ordenadores de la red Internet. En él se basan el resto de protocolos y aplicaciones de comunicación y búsqueda de Internet.

Telnet: un protocolo que permite trabajar a distancia sobre otro ordenador.

Usenet: red mundial de grupos de noticias agrupadas por miles de temas. Para leer las *news* hay que conectarse a un servidor de noticias y usar un programa lector de noticias. Cuando se contesta a una noticia, la respuesta se envía al grupo de *news* y la pueden leer todos los suscriptores.

Web (véase *World Wide Web*).

World Wide Web (WWW): Malla Multimedia Mundial. Un sistema que organiza la información y otros recursos de Internet mediante enlaces de hipertexto y que permite explorar dichos recursos usando un *Browser*.

BIBLIOGRAFÍA GENERAL

Relación de las principales referencias útiles sobre INTERNET:

Caballar JA. Internet. El mundo en tus manos. Madrid: Rama, 1994.

Comer D. The Internet book: Everything you need to know about computer networking and how the Internet works. Englewood Cliffs: Prentice Hall, 1994.

Eddings J. How the Internet works. Emeryville: Ziff-Davis Press, 1994.

Engst AC. Internet Starter Kit for Macintosh (2ª ed.). Indianapolis: Hayden Books, 1994.

Engst A, Low C, Simon M. Internet starter kit for Windows. Indianapolis: Hayden Books, 1994.

Gilster PA. The Internet made easy. *Compuserve Magazine*, junio de 1995.

Hahn H, Stout R. The Internet complete reference. Berkeley: Osborne-McGraw-Hill, 1994.

Internet World Magazine. On Internet '94: an international title and subject guide to electronic journals, newsletter, texts, discussion lists, and other resources on the Internet. Meckler, 1994.

Krol E. The whole Internet user's guide and catalog (2ª ed.). Sebastopol: O'Reilly & Associates, 1994.

Pastor X, Jaureguizar J. Internet en la práctica médica. *Jano* 1995; 49: 101-105.

Thomas BJ. The Internet for scientists and engineers. Bellingham: SPIE Optical Engineering Press, 1995.