



BIBLIOTECA
FAC. DE INFORMÁTICA
U.N.L.P.

UNIVERSIDAD DE BUENOS AIRES

**SISTEMAS DE HALLAZGO Y RECUPERACION DE RECURSOS
EN INTERNET**

Archie, Gopher, Judhead, Kis, Netfind, Prospero, Veronica, Wais, WWW

Tesis presentada en la
Facultad de Ciencias Exactas y Naturales
Departamento de Computación

Por
Cecilia Marcela SARMIENTO
LU 0116/84

Directores:
Lic Javier DIAZ - UNIVERSIDAD DE LA PLATA
Lic. Roberto BEVILACQUA - UNIVERSIDAD DE BUENOS AIRES



Diciembre de 1994



Sistemas de Hallazgo y Recuperación de Recursos en Internet

Tabla de Contenido

<i>RESUMEN</i>	<i>I</i>
<i>ABSTRACT</i>	<i>II</i>
<i>ALCANCES</i>	<i>III</i>
<i>MOTIVACIÓN</i>	<i>VI</i>
<i>ANTECEDENTES</i>	<i>VI</i>
<i>INTRODUCCIÓN</i>	<i>VI</i>
<i>CAPÍTULO I</i>	<i>1</i>
<i>WHOIS - NETFIND - KIS - PROSPERO</i>	<i>1</i>
<i>WHOIS</i>	<i>1</i>
<i>Como acceder a la base de datos Whois</i>	<i>1</i>
<i>Via un cliente</i>	<i>2</i>
<i>Via telnet</i>	<i>4</i>
<i>Via correo electrónico</i>	<i>4</i>
<i>Que información nos ofrece</i>	<i>4</i>
<i>KIS (KNOWBOT INFORMATION SERVER)</i>	<i>5</i>
<i>NETFIND</i>	<i>9</i>
<i>Como funciona</i>	<i>10</i>

Sistemas de Hallazgo y Recuperación de Recursos en Internet

<i>PROSPERO</i>	13
<i>CAPÍTULO II</i>	15
<i>ARCHIE</i>	15
<i>COMO TRABAJA</i>	16
<i>ARQUITECTURA</i>	18
<i>COMO ACCEDER</i>	20
<i>Via telnet</i>	20
<i>Via un Cliente</i>	23
<i>Via correo electrónico</i>	24
<i>CAPÍTULO III</i>	26
<i>GOPHER</i>	26
<i>MODELO DE LOS DATOS</i>	27
<i>INTERFACE GOPHER</i>	31
<i>ARQUITECTURA</i>	32
<i>Protocolo Gopher</i>	33
<i>El cliente</i>	38
<i>El servidor</i>	38
<i>Servidores CSO</i>	39
<i>Servidores de búsqueda de texto completo</i>	40
<i>Conjunto de tipo de items</i>	40
<i>Protocolo Gopher+</i>	41
<i>SERVIDOR GOPHER</i>	42

Sistemas de Hallazgo y Recuperación de Recursos en Internet

<i>COMO ACCEDER</i>	42
<i>Via un cliente</i>	42
<i>Via telnet</i>	44
<i>VERONICA</i>	44
<i>JUGHEAD</i>	45
<i>CAPÍTULO IV</i>	47
<i>WAIS</i>	47
<i>ARQUITECTURA</i>	47
<i>Protocolo Z39.50</i>	49
<i>COMO TRABAJA</i>	50
<i>COMO ACCEDER</i>	51
<i>Via un software cliente</i>	51
<i>Via Gopher</i>	51
<i>Via telnet</i>	51
<i>COMO REALIZAR UNA BÚSQUEDA</i>	52
<i>Encontrar fuentes apropiadas</i>	52
<i>Buscar un item</i>	53
<i>CARACTERISTICAS</i>	55
<i>Como acceder a la información</i>	55
<i>Relevance Feedback</i>	55
<i>Selección de los documentos mas adecuados</i>	56
<i>CAPÍTULO V</i>	58
<i>WORLD WIDE WEB</i>	58

Sistemas de Hallazgo y Recuperación de Recursos en Internet

<i>INTRODUCCIÓN</i>	58
<i>Hipertexto, ¿Qué es?</i>	59
<i>URL: Universal Resource Locator</i>	60
<i>EL MODELO DE DATOS</i>	60
<i>ARQUITECTURA</i>	62
<i>PROTOCOLOS</i>	63
<i>File transfer protocols - FTP</i>	63
<i>Network News Transfer Protocol - NNTP</i>	63
<i>Hipertext Transfer Protocol - HTTP</i>	63
<i>Conexión</i>	64
<i>Requerimiento</i>	64
<i>Respuesta</i>	64
<i>Desconexión</i>	64
<i>COMO ACCEDER</i>	65
<i>Vía software cliente</i>	65
<i>Vía telnet</i>	66
<i>Vía una Interface Gráfica</i>	67
<i>TÉCNICAS DE NAVEGACIÓN Y HERRAMIENTAS</i>	67
<i>Definir una estructura</i>	67
<i>Mecanismo de historia</i>	67
<i>Índices</i>	67
<i>Menú de enlaces</i>	68
<i>Una web de índices</i>	68
<i>La vista del proveedor de información</i>	68

Sistemas de Hallazgo y Recuperación de Recursos en Internet

<i>Servidores.....</i>	<i>70</i>
<i>RELACIONES ENTRE WWW Y LOS DEMÁS SISTEMAS</i>	<i>71</i>
<i>El gateways WWW a archie y ftp.....</i>	<i>71</i>
<i>Wais y Web.....</i>	<i>71</i>
<i>Sistemas de Menús y Web.....</i>	<i>71</i>
CONSIDERACIONES FINALES.....	73
APÉNDICE A.....	79
<i>COMO WWW DENOMINA A LOS DOCUMENTOS.....</i>	<i>79</i>
<i>ESQUEMAS DE NOMBRES DE WWW.....</i>	<i>80</i>
<i>Nombres de subesquemas de nombramiento.....</i>	<i>81</i>
<i>Direccionamiento para una búsqueda indexada.....</i>	<i>82</i>
<i>Direcciones de archivos WWW.....</i>	<i>83</i>
<i>Accesos a directorios</i>	<i>84</i>
<i>Dirección de Hipertexto para News</i>	<i>84</i>
<i>Nombres relativos.....</i>	<i>85</i>
<i>Direccionamiento HTTP.....</i>	<i>86</i>
<i>Direccionamiento Telnet.....</i>	<i>87</i>
<i>Direccionamiento Gopher.....</i>	<i>88</i>
<i>Direccionamiento para servidores Wais.....</i>	<i>89</i>
<i>Sintaxis de direccionamiento WWW: BNF</i>	<i>90</i>
<i>Evitando caracteres ilegales.....</i>	<i>92</i>
APENDICE B.....	94

Sistemas de Hallazgo y Recuperación de Recursos en Internet

<i>LENGUAJE HTML</i>	94
<i>INTRODUCCIÓN</i>	96
<i>Texto estructurado</i>	96
<i>Rónulos</i>	97
<i>Tipo de Elemento</i>	97
<i>vacio</i>	97
<i>Caracter</i>	98
<i>Dato caracter reemplazable</i>	99
<i>Contenido elemento</i>	100
<i>Contenido mixto</i>	101
<i>Comentarios y otras señalizaciones</i>	102
<i>Lineas separadas</i>	102
<i>ELEMENTO HTML</i>	102
<i>Propiedades del documento entero</i>	103
<i>Formateo de texto</i>	103
<i>EJEMPLO</i>	104
<i>GLOSARIO</i>	108
<i>BIBLIOGRAFÍA</i>	117
<i>INDICE</i>	120

Tabla de Figuras

<i>COMO TRABAJA ARCHIE</i>	17
<i>ARQUITECTURA ARCHIE</i>	19
<i>ARQUITECTURA GOPHER</i>	43
<i>ARQUITECTURA WAIS</i>	48
<i>VISTA DEL MUNDO WWW</i>	61
<i>ARQUITECTURA WWW</i>	65
<i>VISTA DEL PROVEEDOR DE INFORMACIÓN</i>	69
<i>ILUSTRACIÓN DE LOS DATOS DE HIPERTEXTO CODIFICADO</i>	95



Resumen

La evolución mundial de las redes de computación tiene su mejor ejemplo en Internet tanto en cobertura geográfica, como en cantidad de nodos, volumen de información disponible, población de usuarios.

Los servicios de la mayoría de estas redes han evolucionado mucho desde los primitivos sistemas de mensajería, conexión remota y transferencia de archivos a sistemas distribuidos de información basados en el modelo cliente-servidor. A pesar de ser relativamente nuevos (surgen a partir de 1991) existen una diversidad de opciones: Archie, Wais, Gopher, WWW, etc..

El presente trabajo permite conocer, seleccionar, acceder, usar y entender principios de la arquitectura de los sistemas más difundidos. Trata de evitar el uso de términos técnicos específicos a pesar que se profundiza en los siguientes aspectos:

v Se cubren en forma exhaustiva las opciones existentes

v Se detalla el uso de cada sistema.

v Se provee los mecanismos internos y la arquitectura.

Por último, la presente tesis permite contar con material bibliográfico actualizado en lenguaje español.



Abstract

The world-wide evolution of the computing networks have their best example in the Internet, not only in the geographical coverage but also in the amount of nodes, available information, volume, users population.

The services that the majority of these networks can provide have greatly improved from the primitive systems of e-mail, remote login and file transfer to distributed information systems based on the client-server model. In spite of the fact that these last mentioned are relatively news systems (emerging in 1991), there exists a variety of options.

This paper allows you to know, select, access, use and understand the principles of architecture of the most well-know systems. It tries to avoid the use of specific technical terminology, however it goes deeply into the following aspect:

- v The different options are shown exhaustively.***
- v The use of each systems is detailed.***
- v The internal mechanism and the architecture are provided.***

Finally, this Thesis provides you with an update spanish bibliography.



Alcance

La presente tesis de Licenciatura establece las bases para una comparación de los distintos sistemas de búsqueda y recuperación de información en Internet.

Se partió de una recopilación de información dispersa utilizando documentos originales, publicaciones en congresos, libros generales y de referencias. Las siguientes fueron las seleccionadas por su actualidad, compatibilidad y porque reunían el nivel de información requerida:

Bob Alberti, Farhad Anklesaria, Paul Linder, Mark Mchill, Daniel Torrey: *The Internet Gopher Protocol - The distributed documents Search and Retrieval Protocol*, Universidad de Minnesota Microcomputer and Workstation Network Center, 1992.

Tim Bernes-Lee, Robert Caillau: "World Wide Web". Paper que se presentó en la conferencia de Computing in High Energy Physic, del 23-27 de septiembre de 1992, Francia.

Tim Bernes-Lee, Robert Cailliau, Jean-Françoise Groff, Bernd Pollermann: "World Wide Web: The Information Universe ", publicado en *Electronic Networking: research, Applications and Policy*, Vol2, No. 1, 1992

Tim Bernes-Lee "The World Wide Web Iniciative", Inet'93 , Proceedings Agosto 1993.

Peter Danzing, Katia Obraczka and Shih-Hao Li: Internet Resource Discovery Services, Computer Science Department University of Southern California

Franklin Davis, Brewster Kahle, Harry Morris, Jim Salem, Tracy Shen, Rod Wang, Jhon Sui, Mark Grinbaum: WAIS Interface Protocol Prototype Functional Specification, Thinking Machine Corporation, Dow Jones & Company, Inc, 1990.

Bennett Falk : "The Internet Roadmap", Sybex, 1994.

Mark Gib & Richard Smith: " Navigating the Internet" Sams Publishing, 1993.

Nancy John & Edward Valauskas: "The Internet Troubleshooter" American Lib. Associations, 1994.

Internet Gopher Server Documentation - University of Minnesota. Documentación que acompaña al servidor Gopher, v.1.1, 1992.

Harley Hahn & Rick Stout: The Internet Complete Reference, MacGrawHill, 1994.

Brewster Kahle: Wide Area Information Server Concepts, Thinking Machine Corporation.

Brenda Kehoe: Zen and the Art of Internet A beginner's Guide, Prince Hall, 3ra. edición, 1994.

Jonathan Kochmer and NorthWestNet: Internet Passport, NorthWestNet, 1993.

Ed Krol, The Whole Internet user Guide & Catalog, O'Reilly & Associates, 1994, 2da. edición.

Tracy Laquey with Jeanne Ryer: The Internet companion, Addison Wesley Publishing Company, 1993.

Tim Bernes-Lee, "Word Wide Web Book" , 1992.

**Tim Bernes-Lee, Daniel Connolly : "Hypertext Markup Language:
A representation of textual Information and
Metainformation for Retrieval and Interchange ", 13 de
Julio de 1993.**

**April Marin: Getting Started, Internet Information Series,
1992.**

**User Glossary Working Group:RFC 1392, Internet Glossary,
1993.**

**Michael Schwartz, Alan Emtage, Brewster Kahle, Cliffrod
Neuman: A Comparison of Internet Resource Discovery
Approaches, Agosto 1992.**

**Petter Otte: " The Information Superhigway beyong the
Internet", 1994.**

Ansi Z39.50 Version 2 THIRD DRAFT, Mayo 1991.

**Cricket Lin, Jerry Peek, Russ Jones, Brian Buns & Adrian Nye:
"Managing Internet Information Services", O'Reilly &
Asociations, 1994.**

**Eric Lease Morgan: "Wais & Gopher Servers, A guide for Internet
End-User", Meck Lermedio, 1994.**

**Paul Gilster: "Finding it on the Internet, The Essential guide",
John Wiley & Sons, Inc, 1994.**

**A la información relevada se la seleccionó y se la estructuró en
un formato uniforme, con un mismo estilo, a fin de simplificar la
lectura.**

Motivación



La gran difusión de Internet como plataforma de experimentación de la autopista de información promovida por el Vicepresidente de los Estados Unidos Al Gore, ha causado una gran avalancha de material informativo con escaso fundamento técnico y con relevamientos muy incompletos.

El presente trabajo trata de cubrir este bache presentado las herramientas de uso mas difundidas en Internet y dando una profusa bibliografía donde recurrir para profundizar en cada tema.

Se trabajó completando el panorama de las herramientas disponibles en redes de computación utilizando un enfoque dual, desde el punto de vista de las posibilidades para el usuario como desde el punto de vista de las estructuras en que se basan los servicios. Esta tesis busca ser completa en cuanto a la cantidad de herramientas analizadas y para cada una, se profundiza al máximo nivel posible sin entrar en detalles de implementación.

Antecedentes

No existen antecedentes en la Argentina de trabajos similares.

Introducción

Es indudable que hoy en día Internet, con más de 40.000 redes interconectadas alrededor del mundo, es la plataforma sobre la cual se está edificando la Infraestructura Global de Información.

Si bien desde sus comienzos Internet ha servido como una red para intercambio de información, es recién en los últimos tres años que han comenzado a desarrollarse e implementarse herramientas que permitan una mejor y eficiente organización, búsqueda y acceso al creciente volumen de información disponible en línea a través de la red.

En un primer momento el auge fueron los servidores de ftp anónimo, por medio de ellos se lograba acceder a un importante volumen de información utilizando simplemente el servicio de transferencia de archivos, uno de los tres servicios básicos de Internet.

Con el aumento creciente de servidores e información, algunos grupos comenzaron voluntariamente a compilar listas de servidores de ftp anónimo con alguna mención a la información disponible en cada uno de ellos, estas listas se distribuían por medio de correo electrónico, y a través de diversos grupos de USENET News, y obviamente también estaban accesibles por ftp anónimo en sitios prefijados.

Al principio estas listas resultaron ser de gran ayuda dado que permitían conocer la existencia de los servidores de ftp anónimo disponibles, pero no constituían un mecanismo eficiente para buscar información, dado que no era posible incluir en la lista el contenido completo de cada servidor archivo por archivo, además la cantidad y variedad de información en los servidores crecía día a día, por lo que la actualización y consistencia de la lista cada vez se tornaba más engorrosa.

No pasó mucho tiempo hasta que un grupo de la universidad de McGill en Canadá desarrollara e implementara una solución sistemática a este problema, creando una herramienta que se conoce como ARCHIE, que veremos en detalle más adelante y que aún se encuentra en uso en Internet.

Paralelamente, varios grupos comenzaron a estudiar distintas formas de encarar el problema de cómo catalogar y organizar la información, y en la creación de protocolos y herramientas más eficientes para la búsqueda y acceso a la información. Comienza entonces a tomarse a los Sistemas de Hallazgo y Recuperación de la Información (más general y conocido en Internet como Resource Discovery) como un tema de estudio de alto interés en Internet.

Estos grupos dieron como resultado un conjunto de novedosas herramientas como GOPHER, WAIS, WWW, etc. Muchas de estas se han basado en la modalidad cliente-servidor, existiendo software de uso público e implementaciones comerciales, tanto para los clientes como para los servidores, para una gran cantidad de plataformas de hardware y sistemas operativos.

Ha sido notoria la gran aceptación de estas nuevas herramientas, un claro ejemplo es la WWW o World Wide Web basada en el protocolo de transferencia de hipertexto o HTTP, tan solo en dos años la cantidad de servidores WWW en el mundo ha crecido desde unas pocas decenas a más de 4.000, distribuidos en los lugares más recónditos del planeta y con una amplia variedad de información.

En particular, los servidores de WWW y los clientes asociados como distintas variedades del NCSA MOSAIC, están produciendo una revolución en Internet. Son varios los motivos, entre ellos la disponibilidad de interfaces de usuario con capacidad gráfica y de fácil utilización (point-and-click), la posibilidad de "conectar" distintos documentos u otro tipo de información relacionada con un mismo tema o ideas por medio del hipertexto, con la incorporación, aún limitada, de facilidades de multimedia, el acceso a través de una única interfaz de usuario a otros servicios de Internet y la generación de nuevas aplicaciones interactivas utilizando el mismo protocolo de comunicación e interfaz de usuario,. etc.

Estas y otras características de la WWW, han determinado que sea la herramienta favorita para quienes desean proveer servicios de

información, particularmente para las empresas que quieren dar a conocer sus productos y servicios a través de Internet, así como para muchos tiempo atrás disponer de una dirección de correo electrónico e incluirla en su tarjetas de negocios representaba un detalle de estilo y pertenencia, hoy sucede lo mismo con WWW, la mayoría de las grandes corporaciones ya poseen su servidor WWW y hacen referencia a él para obtener más información de la empresa, productos y servicios, es más, ya existen servidores que poseen un servicio de orden de compra en línea, donde el usuario luego de obtener información de un producto determinado puede completar la orden de compra en el momento en forma interactiva.

Si bien WWW presenta muchas posibilidades que solo tienen como límite la imaginación, desde el punto de vista del tema a tratar, "Resource Discovery", todavía resta mucho por hacer, tanto en investigación como desarrollo de nuevas herramientas, pero es claro que el desarrollo de la Infraestructura Global de Información está en marcha y es necesario participar en el mismo.

Argentina no está ajena a este desarrollo, con los avances producidos los últimos tiempos en conectividad nacional e internacional de las distintas redes que conforman la red Internet Argentina, ya han comenzado a aparecer varios sitios que proveen información en línea a través de Internet, sin embargo el grado de desarrollo comparado con otros países es aún muy bajo.

Resulta entonces necesario para participar en forma activa del desarrollo de la Infraestructura Global de Información tener conocimiento y adquirir experiencia local en la implementación de estos servicios.

Si bien el tema es muy amplio, este trabajo presentará varios aspectos de las herramientas más tradicionales y de mayor prestación a los usuarios, con el objetivo de proveer un panorama general sobre cada una de estas citando las referencias bibliográficas correspondientes para quien desee profundizar en un tema en particular.

Dado que no es sencillo caracterizar de forma única a la gran variedad de información disponible, la definiremos en forma genérica como un recurso.

Un recurso es todo aquello de interés que puede ser accedido por el usuario. Como por ejemplo: hardware (supercomputadores, impresoras, etc.); un grupo de personas que contestan o resuelven un problema; listas de correo electrónico; sistemas de conferencias, etc. Información en línea es otro conjunto de recursos disponible bajo demanda, como por ejemplo, bases de datos, documentos, software, gráficos, imágenes, catálogos de bibliotecas, boletines informativos, archivos de audio, etc.

Toda esta información puede ser provista por cualquier usuario que quiera poner datos accesibles a la comunidad, ya sea en forma gratuita o no. Debido a esto, la cantidad de información es inmensa así como también su complejidad, debido a que no existen normas sobre como organizarla y almacenarla.

Con el objetivo de ayudar al usuario a buscar y recuperar información en este espacio de millones de gigabytes alrededor del mundo, grupos de investigación desarrollaron distintos sistemas que simplifican el acceso a estos recursos, los cuales veremos a través de este trabajo.

Estos sistemas se pueden agrupar según su función en: Sistemas para obtener información sobre usuarios de Internet, como Whois, Netfind, Kis; Sistemas para organizar la información del usuario como Prospero; y Sistemas para buscar recursos en general como por ejemplo, Archie, Gopher, WAIS y WWW.

En el Capítulo I veremos algunos ejemplos de herramientas para la búsqueda de información sobre usuarios . Elegí las más populares como por ejemplo Whois, KIS, Netfind. También podemos encontrar en este capítulo una herramienta para organizar como lo es Prospero. En los capítulos que siguen veremos mas detalladamente los sistemas Archie en el capítulo I. En el Capítulo

III Gopher, Veronica y Jughead. En el Capítulo IV WAIS. Y finalmente en el Capítulo V WWW. El trabajo se completa con una serie de Apéndices con mayor información y descripción técnica de algunos elementos componentes de WWW además de contar con un glosario.

Me gustaría aclarar que estas no son las únicas herramientas que existen actualmente en el Internet. Cada día aparece alguna nueva, o tal vez una actualización de las que ya están en uso. Podemos percibir un cambio en cada momento y es tan vertiginoso que es imposible tener los datos actualizados al momento por lo que podemos encontrar desactualizaciones en este trabajo.



CAPITULO I

Whois - Netfind - Kís - Prospero.

El correo electrónico es hoy en día uno de los recursos mas utilizados en el Internet. La mayoría de los programas de correo son muy faciles de utilizar pero podemos encontrarnos con un obstaculo, el desconocimiento de la dirección a donde enviar nuestro mensaje.

Una posible solución es utilizar algunas de las herramientas que fueron diseñadas con este proposito. Veremos en este capítulo tres de las mas utilizadas para encontrar direcciones de usuarios, como por ejemplo, Whois, Netfind y Knowbots.

Para finalizar veremos una herramienta que nos ayuda a organizar virtualmente datos llamada Prospero.

WHOIS

Whois es el nombre que se le da tanto a la base de datos conteniendo información sobre nombres de usuarios de Internet con sus direcciones electrónicas, como al comando que permite buscar en ella.[KRO/94]

Este directorio es mantenido por el Centro de Información de red (NIC) y contiene alrededor de 70.000 entradas.[Gil/94] La Forma en que estan almacenados los datos es:

Apellido, Nombre, Información variada.[KRO/94]

Como acceder a la base de datos Whois:

Existen tres formas directas de acceder a este servicio en forma directa, mediante un cliente, realizando un telnet o a traves del correo electrónico. Existen otras formas pero estas son a traves de otros servicios como por ejemplo Gopher . Veamos las formas directas.

□ **Whois via un cliente:**

Veamos que es el modelo cliente/servidor. Este modelo proporciona un mecanismo para que dos computadoras cooperen en una misma tarea. Este intercambio se logra mediante dos programas uno llamado cliente y el otro llamado servidor, y una forma de entendimiento entre ambos llamado protocolo.[Mor/94]

Las tareas básicas del cliente son tres. Una es mostrarle al usuario los datos que el cliente consigue del servidor. Segundo, iniciar un proceso de comunicación con el servidor. Y por último interpretar la información enviada por este.[Mor/94]

El servidor tiene como tareas básicas, primero interpretar la comunicación del cliente. Segundo, realizar el procesamiento necesario para obtener lo pedido. Tercero, devolverle los datos resultantes al cliente para que este haga su trabajo. [Mor/94]

La principal ventaja de este modelo es que gracias a esta independencia en los programas, se puede desarrollar una interface usuaria independiente de como sea la forma de almacenamiento o administración de los datos. Por ejemplo, podemos mostrarle al usuario una pantalla con graficos, color y sonido a pesar que los datos esten almacenados en una computadora Unix. Podemos utilizar una computadora con un gran poder visual para el cliente y una con gran poder en procesamiento y almacenamiento para el servidor, mejorando notablemente la performance del sistema. [Mor/94]

La manera más fácil es correr el comando whois en la propia computadora y el cual llamará al programa cliente. La forma de usar este comando es:

% whois <apellido>

busca en el servidor definido como default.

% whois '<apellido>,<nombre>'

busca en el servidor definido como default. De esta forma podemos restringir la búsqueda si sabemos el nombre completo.

% whois -h <nombre-servidor-whois> <apellido>

busca en el servidor nombre-servidor-whois.

Veamos un ejemplo:

% whois -h ds.internic.net Jorge

La respuesta será:

The ds.internic.net whois server is being queried:

**Garza, Jorge (JG165) Garza@MCC.COM
Microelectronics and Computer Technology Corporation
3500 West Balcones Center Drive
Austin, TX 78759-6509
(512) 343-0978**

Record last updated on 19-Nov-90.

**Beauchamp, Jorge L. Rodriguez (JLRB) SYSTEM@NAIC.EDU
National Astronomy and Ionosphere Center
Arecibo Observatory
P.O. Box 995
Arecibo, PR 00613
(809) 878-2612**

Record last updated on 09-Feb-89.

**Seidel, Jorge T. (JTS15) seidel@AEROSPACE.AERO.ORG
The Aerospace Corporation
Mail Stop M1/099
P.O. Box 92957
Los Angeles, CA 90009-2957**

Record last updated on 10-Jul-91.

**Rufat-Latre, Jorge (JR43) jrufat@ATHENA.MIT.EDU
Massachusetts Institute of Technology
305 Memorial Drive
Cambridge, MA 02139
(617) 225-9826**

Record last updated on 23-Mar-92.

The rs.internic.net whois server is being queried:

**Fernades, Jorge Cabral (JCF) jorge@BRUFRN.BITNET +55-84-231-8242
Gomes, Jorge (JG378) jorge@UXLIP1.LIP.RCCN.PT +351 773880
The nic.ddn.mil whois server is being queried:
Gomes, Jorge (JG378) jorge@UXLIP1.LIP.RCCN.PT +351 773880
Jorge, Kevin (KJ95) ASQK-J-RU@SEOUL-EMH1.ARMY.MIL (DSN) 722-4317**

□ **Whois via telnet:**

Tenemos la facilidad de poder ingresar a un servidor de acceso público utilizando el comando telnet con el nombre de usuario whois.

telnet <nombre_servidor_whois>

Como por ejemplo podemos intentar acceder a Whois con,

telnet whois.ncsu

Las búsquedas pueden ser por dominio, por nombre de computadora, en un sitio específico o en un área dada por el usuario, por nombre de usuario, por parte de la dirección de correo (usuario@ o @maquina)

□ **Whois via correo electrónico.**

Existe un servicio automático que permite a los usuarios acceder al whois mediante correo electrónico. Es especialmente útil para aquellas personas que no tiene acceso directo a Internet como por ejemplo sitios BITNET y UUCP.

Para usar este servicio solo hay que enviar un mensaje a service@nic.ddn.mil y colocar en el sujet: WHOIS <nombre>. Sin embargo si el subject esta vacío se procesara la primera línea del cuerpo del mail. Para obtener información de como usar este servicio se puede colocar el comando WHOIS HELP en el subject.

Que información nos ofrece:[SCH/92]

La información que nos ofrece es por ejemplo:

- **nombre de la persona**
- **dirección del trabajo**
- **número de teléfono**
- **dirección electrónica**

Como cada servidor WHOIS colecciona información geográficamente distribuida, en una única base de datos, provee un buen punto focal de registración y búsquedas. Sin embargo, cada servidor contiene solo una pequeña fracción de los usuarios de internet y sitios que han sido registrados en el NICs. Contiene además, información desactualizada debido a que los usuarios olvidan de informar cuando ocurren cambios. Más aun como cada servidor Whois corre independiente uno del otro (sin coordinación en el contenido o formato) los usuarios deben tratar con la distribución e inconsistencia entre los servidores.

Otro de los problemas es que Internet continua creciendo. En estas circunstancias, un directorio centralizado comenzara a ser un cuello de botella y un punto critico de falla.

KIS (Knowbot Information Server)[KOC/93]

Knowbots es una aplicación que conduce el requerimiento del usuario a varios sitios en Internet. En cada sitio el requerimiento es implementado en comandos específicos según sea el servicio ofrecido en ese lugar.

Kis recupera información acerca de usuarios en Internet. Esta información abarca direcciones de e-mail, números telefónicos, direcciones postales y datos varios sobre la institución. Permite buscar a través de varios servicios de directorio, como whois (nic.ddn.mil), finger, entre otros. Existe un cliente para el sistema operativo Unix.

Las formas de acceso son mediante un cliente, correo electrónico o realizando un telnet a algun servidor kis.

Veamos un ejemplo de una búsqueda al servidor público de Kis mediante :

```
% telnet info.cnri.reston.va.us 185
```

```
Trying 132.151.1.15...
Connected to info.cnri.reston.va.us.
Escape character is '^]'.
      Knowbot Information Service
KIS Client (V2.0).  Copyright CNRI 1990.  All Rights Reserved.
KIS searches various Internet directory services to find someone's street address,
email address and phone number.

Type 'man' at the prompt for a complete reference with examples.
Type 'help' for a quick reference to commands.
Type 'news' for information about recent changes.

Backspace characters are ' ' or DEL
Please enter your email address in our guest book...
(Your email address?) > ceci@secyt.gov.ar
> query cecilia sarmiento
Trying whois at ds.internic.net...
The ds.internic.net whois server is being queried:
No match for "SARMIENTO and CECILIA"

The rs.internic.net whois server is being queried:
Sarmiento, Cecilia Marcela (CMS4)          ceci@SECYT.GOV.AR
  Secretaria de Ciencia y Tecnologia
  Red Cientifica y Tecnologica Nacional
  Cordoba 831 3er Piso
  (1054) Buenos Aires
  AR
  +54-1-312-8917

The nic.ddn.mil whois server is being queried:
No match for name "SARMIENTO,CECILIA".
Trying mcimail at cnri.reston.va.us...
(Press RETURN to continue)
Trying ripe at whois.ripe.net...
Trying whois at whois.lac.net...
Sarmiento, Cecilia M. (CMS) ceci@secyt.gov.ar
  Secretaria de Ciencia y Tecnologia
  Red Cientifica y Tecnologica Nacional (RECYT)
  Av. Cordoba 831 - 3er. piso
  Capital Federal, 1054
  Republica Argentina
  Tel: +54-1-3128917
  Fax: +54-1-3138389
  Ultima Act.(Last Update): Tue 02/01/94 13:06:03
>bye
KIS exiting
```

Veamos el help en línea ofrecido por el servidor:

```
> ?
KIS (Telnet) Handy Reference:
  (use the command "man" for more complete information and examples)

  (Use one command to a line,
   Use as many lines as you like
   If there is no keyword, we will assume it is a name... )

  These signals should work at any time:
    INTR (usually ^C) will interrupt the current command,
    QUIT (usually ^\) will exit KIS entirely.

  These are the commands:
  -----

  service SERVICE_NAME [SERVICE_NAME ...]
  service NUMBER [NUMBER ...]
    -- Add the specified service to the list of resources to search
    number should come from the menu list that the "services"
    command generates.
(Press RETURN to continue)
  services
    -- Lists the services KIS knows.
  org ORGANIZATION
  country COUNTRY
    -- Restrict the search to persons that work in a specific
    organization or live in a specific country.
    If there is no argument, this will turn off any
    setting you made earlier.
  echo on | off
    -- If on, KIS shows you exactly what commands the server is executing. By
  default this is off.
  ident service-specific-identifier
    -- Some services assign keywords (handles) to individuals in their database.
  This causes a search for a handle.
    If you leave the argument blank, it will turn the handle
    choice off.
  print
    -- Shows the current name, organization, services, etc.
    that you have chosen to this point.
  query USERNAME
    -- Search for someone with the name (or email address) given here.
    The name may contain a first, middle and last name. See the
    examples at the end of this note.
  help (or ?)
```

-- Print this handy reference.
man
-- Print the manual page entry.
news
-- Print version info and other useful facts about KIS.
set lines [N]
-- With integer argument 'N':
 Display N lines of text before prompting for a carriage return.
-- With integer argument 'O':
 Display all lines of text.
-- Without the integer argument:
 Print the number of lines now being displayed before a prompt
unset lines
-- Reset the number of lines displayed to the default (20).
gripe
-- Log complaints or compliments about the KIS service
quit
exit
-- Exeunt.

Note: Please type the entire command word as you see it in the list.

Kis does not understand abbreviations.

Arguments are represented here by CAPITAL letters, but you may type them in small letters too

[Arguments] in square brackets are optional

Arguments ... followed by dots indicate that you may use several arguments in the same command.

A vertical the bar "|" indicates that you may pick one option.

A word on user names:

Anything similar to these will work fine---

Jones John Jones Jones, John Jones, J.

John..., John

-- will match John Johnson, John Johnsonovich, John Johns, etc. Avoid titles (Dr. John Jones, Jr.): we can't handle those yet.

Netfind [SCH/92]

Es un servicio de directorio que intenta localizar direcciones electrónicas y otro tipo de información sobre usuarios de Internet. Utiliza un conjunto de heurísticas para localizar los host sobre los cuales el usuario buscado podrían tener cuenta o casilla de correo electrónica.

Este programa no solo busca el nombre y la dirección de correo electrónico sino que también tratará de encontrar información provista por el comando finger. [HAH/94]

Existe dos formas de usar el Netfind]:

1) Vía un cliente.

El cliente se invoca a través del comando netfind. Solo esta disponible para sistemas Unix.[Gil/94]

2) Via telnet[HAH/94]: se puede utilizar servidores públicos localizados en:

**archie.au (AARNet, Melbourne, Australia)
bruno.cs.colorado.edu (University of Colorado, Boulder)
dino.conicit.ve (Nat. Council for Techn.&Scien.Research, Venezuela)
ds.internic.net(InterNIC Directory & DB Services, S.Plainfield, NJ)
eis.calstate.edu (California State University, Fullerton, CA)
krnic.net (Korea Network Information Center, Taejon, Korea)
lincoln.technet.sg (Technet Unit, Singapore)
malloco.ing.puc.cl (Catholic University of Chile, Santiago)
monolith.cc.ic.ac.uk (Imperial College, London, England)
mudhoney.micro.umn.edu (University of Minnesota, Minneapolis)
netfind.fnet.fr (Association FNET, Le Kremlin-Bicetre, France)
netfind.icm.edu.pl (Warsaw University, Warsaw, Poland)
netfind.if.usp.br (University of Sao Paulo, Sao Paulo, Brazil)
netfind.sjsu.edu (San Jose State University, San Jose, California)
netfind.valib.cz (Liberec University of Technology, Czech Republic)
nic.uakom.sk (Academy of Sciences, Banska Bystrica, Slovakia)
redmont.cis.uab.edu (University of Alabama at Birmingham)
macs.ee.mcgill.ca. (McGill University Netfind server)**

Ejemplo de servidores públicos Netfind.

Como Funciona:

i. Se realiza la consulta

El usuario de Netfind especifica que persona busca con :

- **nombre, apellido o login name**
- **una o más claves describiendo el nombre o ubicación de donde el usuario trabaja**

ii. Las claves se usan para buscar en un Base de Datos SEED y así obtener pistas de dominios administrativos potenciales donde buscar. Esta base de datos es actualizada mediante información obtenida monitoreando distintas fuentes de datos incluyendo:

- **mensajes USENET en los electronic bulletin board,**
- **dominios WHOIS de distintos centros de información de red.**
- **logs de distintos servicios de red**
- **información suministrada por los usuarios**

iii. Basado en las conclusiones obtenidas de la consulta a la base de datos anterior, el usuario es consultado para que seleccione un subconjunto de los dominios a buscar.

iv. Netfind busca esos dominios en paralelo de la siguiente forma:

I. cada dominio es buscado en el Domain name systems DNS, para localizar nombres de los servidores del dominio. Estos a menudo corren sobre máquinas centrales para administrar con cuentas e información de mail para muchos usuarios en ese lugar.

II. Consulta a los servidores de Simple Mail Transfer Protocol (SMTP) sobre las máquinas donde el servidor de nombres corre con el objetivo de encontrar información de mail sobre el usuario.

III. Si la información es encontrada, se consulta a la máquina mediante el servicio FINGER.

Veamos un ejemplo haciendo telnet a macs.ee.mcgill.ca.

```
Trying 132.206.69.1...
Connected to macs.ee.mcgill.ca.
Escape character is '^]'.

SunOS UNIX (finnegan)

login: netfind
Password: <enter>

=====
Welcome to the McGill University Netfind server
=====

Alternate Netfind servers:
  archie.au (AARNet, Melbourne, Australia)
  ...
  redmont.cis.uab.edu (University of Alabama at Birmingham)

I think that your terminal can display 24 lines.  If this is wrong, please enter the "Options"
menu and set the correct number of lines.
Top level choices:
  1. Help
  2. Search
  3. Seed database lookup
  4. Options
  5. Quit (exit server)
--> 2
Enter person and keys (blank to exit) --> sarmiento secyt gov ar
Searching secyt.gov.ar
( 1) got nameserver lanin.secyt.gov.ar
( 1) got nameserver proteus.secyt.gov.ar
( 1) SMTP_Finger_Search: checking domain secyt.gov.ar
Found the mailbox 'sarmiento@secyt.gov.ar', but this may not be user's home mailbox.
( 1) get_host_addr: gethostbyname for secyt.gov.ar failed
SYSTEM: secyt.gov.ar
  Login name: ceci                In real life: Cecilia Sarmiento
  Office: RECYT-NOC, +541-312-8917
  Directory: /usr/ceci            Shell: /bin/csh
  On since Jan 8 10:33:49 on ttyS0b      14 seconds Idle Time
  No Plan.

  Login name: chuli                In real life: Cecilia M. Sarmiento -RECYT - Via UUCP
  Directory: /usr/spool/uucppublic      Shell: /usr/lib/uucp/uucico
  Last login Sun Jan 8 06:09
  No Plan.

Unable to parse line 'Last login Sun Jan 8 06:09'
( 1) SMTP_Finger_Search: secyt.gov.ar leads us to 0 other machines

FINGER SUMMARY:
```

- Found multiple matches for "sarmiento" in finger output, so unable to determine most recent/last login information or most promising electronic mail information. Please look at the above finger search history and decide for yourself which is best.

Continue the search ([n]/y) ? --> y

Domain search completed. Proceeding to host search.

- (1) SMTP_Finger_Search: checking host secyt.secyt.gov.ar
 - (2) SMTP_Finger_Search: checking host gopher.secyt.gov.ar
 - (3) SMTP_Finger_Search: checking host www.secyt.gov.ar
- SYSTEM: secyt.secyt.gov.ar

 Login name: ceci In real life: Cecilia Sarmiento
 Office: RECYT-NOC, +541-312-8917
 Directory: /usr/ceci Shell: /bin/csh
 On since Jan 8 10:33:49 on ttyS0b
 No Plan.

 Login name: chuli In real life: Cecilia M. Sarmiento -RECYT - Via UUCP
 Directory: /usr/spool/uucppublic Shell: /usr/lib/uucp/uucico
 Last login Sun Jan 8 06:09
 No Plan.

Unable to parse line 'Last login Sun Jan 8 06:09'

- (2) connect timed out
- (3) connect timed out

FINGER SUMMARY:

- Found multiple matches for "sarmiento" in finger output, so unable to determine most recent/last login information or most promising electronic mail information. Please look at the above finger search history and decide for yourself which is best.

Continue the search ([n]/y) ? --> y

FINGER SUMMARY:

- Found multiple matches for "sarmiento" in finger output, so unable to determine most recent/last login information or most promising electronic mail information. Please look at the above finger search history and decide for yourself which is best.

Enter person and keys (blank to exit) -->

Top level choices:

- 1. Help
- 2. Search
- 3. Seed database lookup
- 4. Options
- 5. Quit (exit server)

--> 5

Exiting Netfind server...

Prospero[DAN]

Es una herramienta para organizar información distribuida. Prospero permite a los usuarios construir vistas estandarizadas (sistemas virtuales) de directorios distribuidos a través de Internet.

El espacio de nombres forma un grafo dirigido, donde los nodos intermedios son directorios, las hojas son archivos y las flechas son enlaces del sistema. Igual que los sistemas de archivos distribuidos, subárboles de Prospero pueden ser almacenados en distintos servidores de este. Un espacio de nombre de usuario es un subgrafo comenzando con un nodo en particular, que es el root del espacio de ese usuario.

Los usuarios organizan su espacio de nombres jerárquicamente construyendo vistas que son esencialmente directorios compuestos de distintas fuentes que incluyen vistas propias como importadas de otros usuarios. Estos directorios pueden residir en diferentes servidores Prospero. Un tipo especial de vista es un índice, que retorna un directorio de objetos que satisfacen alguna consulta. esto permite a los usuarios de Prospero acceder a otro tipo de búsquedas, como por ejemplo la interface Prospero-Archie, permite a los usuarios construir vistas conteniendo directorios cuyos objetos son el resultado de consultas a Archie.

Los usuarios pueden encontrar la información navegando a través de vistas disponibles. El cliente de Prospero provee a los usuarios con herramientas para navegar que son análogas a las provistas por sistemas de archivos convencionales. El usuario comienza usando el comando vsetup <nombre_del_sistema_virtual> que ubica al usuario en un punto específico del espacio de nombres de Prospero. Luego a través de:

- vcd el usuario cambia de directorio virtual**
- vwd puede ver el directorio actual**
- vls puede ver también el contenido del directorio.**

Cuando el usuario encuentra el objeto de interés, que puede ser un archivo o una vista, este puede incluirlo enlazándolo. El cliente Prospero provee comandos para agregar y borrar enlaces de un nodo actual al espacio de

nombres del usuario. Un enlace, especifica el nombre del host donde el objeto esta almacenado y el nombre local de este. Si el destino de un enlace es un directorio, el lazo provee información para resolver un nombre en aquel directorio. Esta información se obtiene consultando al servidor correspondiente. Para archivos, los lazos contactan al servidor apropiado para proveer la información para el modo de acceso. Un enlace puede incluir datos adicionales como por ejemplo, tipo de enlace y filtros asociados con el. Enlaces especiales permiten que el contenido de un directorio destino sea virtualmente incluido en el directorio físico conteniendo el enlace. Asociando filtros con enlaces, el usuario puede construir vistas estandarizadas de alguna ya existente. Un filtro estandariza la vista destino reorganizando o extrayendo parte de esta.

Los usuarios publican su información registrando su sistema virtual con el administrador del servidor de Prospero. El administrador crea enlaces, al nuevo sistema virtual, en la vista maestra de los sistemas virtuales, donde otros usuarios pueden ver, navegar y si es accesible, enlazar a parte de este.

Existen alrededor de 50 servidores Prospero. Muchos de los usuarios de más de 10.000 sistemas en 30 países corren solamente clientes Prospero. Ellos corren la interface de Prospero a Archie o un cliente completamente Prospero.

CAPITULO II

ARCHIE

El Internet surge como una necesidad de transferencia de información entre centros de investigación. Este movimiento de datos rápido y fácil, comienza a tener una gran aceptación. Rapidamente, esta masa de información distribuida alrededor del mundo plantea un nuevo problema, como hallar el recurso es este espacio de información.

Este espacio de información estaba originalmente comprendido por simples archivos (textos, binarios) desde nuestro punto de vista actual donde el movimiento, las imagenes y el sonido forman parte de nuestras vidas.

En el año 1991, en la Universidad de Mc Gill, Canadá aparece una solución a este problema, Archie.

Archie es una herramienta que nos ayuda en la búsqueda de archivos. Es una solución simple a un problema básico y como tal se mueve dentro de un mundo restringido, el mundo de los sitios FTP anónimos. La ventaja primordial de este mundo restringido es que ofrece información a cualquier usuario sin necesidad de permisos especiales.

Originalmente el propósito fue ahorrar dinero en la utilización de software de dominio público. Con este objetivo, un grupo de personas de la Universidad de Mc Gill comienza a buscar, en forma manual y en los lugares FTP anónimos, que infomación habia disponible. Luego este proceso de búsqueda fue automatizado y gracias a esto nace Archie.

Es interesante como surge el nombre de Archie. Mucho tiempo atrás existía una tira cómica con este nombre. Muchos creían que de ahí surgió el nombre pero no fue así. Antiguamente los comandos en los sistemas Unix tendían a ser cortos por lo que la palabra inglesa archive (archivo) se acortó y formó archie. Rapidamente esta herramienta se hace popular y los usuarios se dirigían a ella como si fuera una persona por lo que se la asoció con el famoso personaje de la tira cómica que lleva su nombre.

Para el año 92, aparece la versión 3.0 y tiene la particularidad de ser comercial. Por esta razón, esta herramienta de búsqueda es ofrecida con upgrate y soporte técnico. Aunque halla un costo para esta versión existen otras que se ofrecen gratuitamente.

Veamos como fue el crecimiento de la cantidad de servidores a través de los años. En el año 1992 hubo más de 20 servidores de acceso público alrededor del mundo accesibles a través de Internet. Por el año 1993 aumento a 27 servidores. En el año 1994 existen 52 servidores pero no todos son públicos; la cantidad de sitios FTP monitoreados son 900, los que ofrecen en conjunto, una cantidad de 3.5 millones de archivos (muchos gigabytes de información).

La popularidad de Archie creció tan abruptamente a punto tal que otras herramientas como Gopher y WWW, que veremos en capítulos posteriores, permiten el acceso a Archie a través de ellas.

I. Como Trabaja

El servicio actual de Archie permite consultar dos bases de datos:

a). La base de datos con todos los nombres de los archivos almacenados en los sitios públicos. El usuario puede buscar en esta base usando, un programa cliente, ingresando remotamente en un servidor Archie, usando Gopher o enviando un mensaje electrónico con comandos apropiados.[DAN]

Para mantener la base de datos actualizada, cada mes el programa Archie realiza un ftp anónimo a todos los sitios contenidos en su lista maestra. Para cada lugar, el sistema colecciona un listado completo del directorio de los sitios ftp anónimos (usando un comando de listado de directorio recursivo "ls -lR) y la agrega a la base de datos central. [KOC/93]

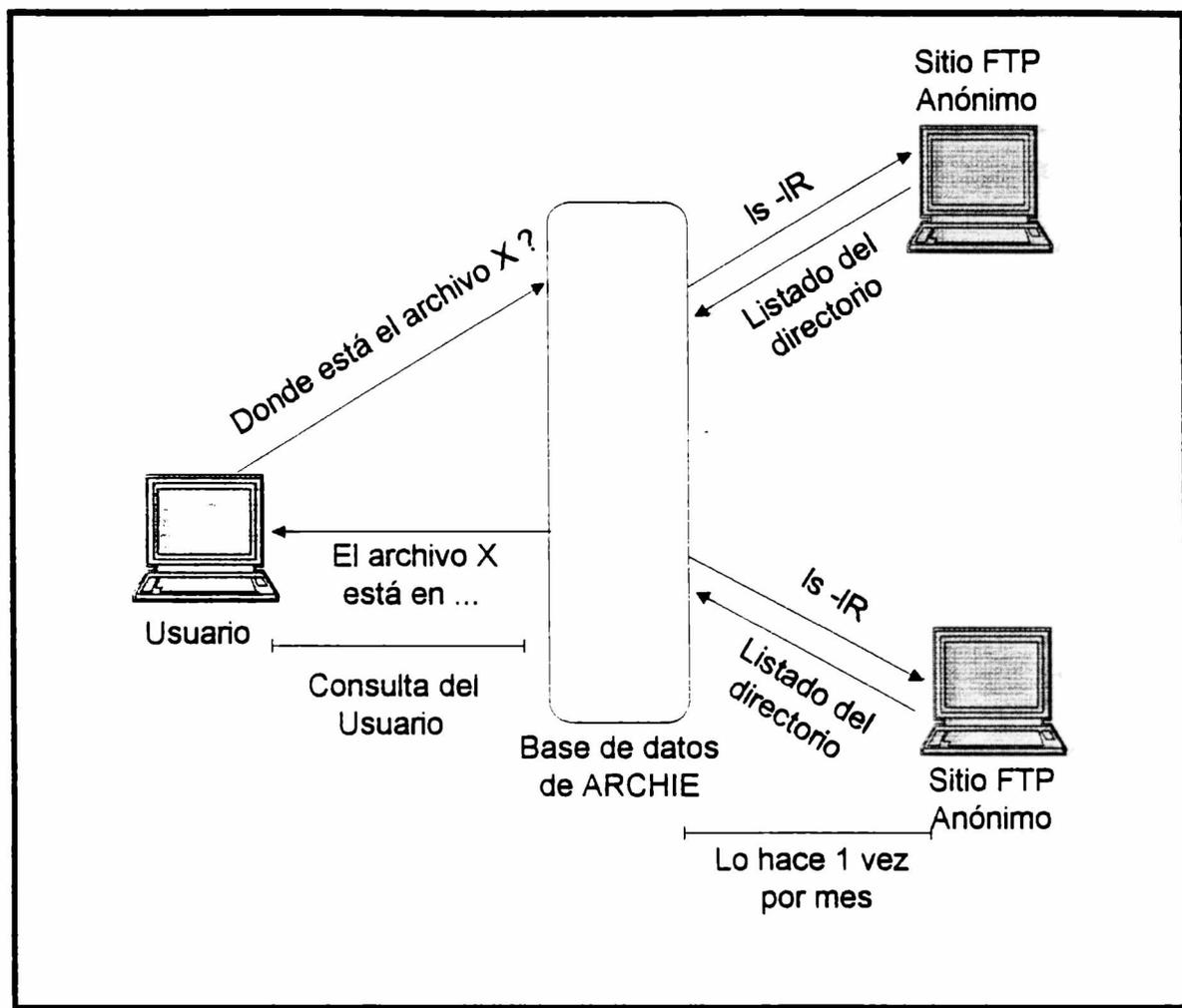


Fig A: Como trabaja Archie [KRO/94]

Se puede consultar esta base de datos a través de distintas interfaces de cualquiera de los servidores duplicados que hay en el mundo, usando expresiones regulares y otros tipo de consultas.

Como resultado de una búsqueda podemos obtener el Internet Domain Name del host, el directorio donde esta ubicado el archivo, el tamaño del archivo en bytes, y la fecha en que el archivo fue almacenado.

Podemos ver que existen 2 pasos para obtener un archivo:

1. hallar el archivo deseado: utilizando Archie

2. obtener una copia del archivo: una vez localizado la información deseada, utilizar las herramientas tradicionales de networking (como por ejemplo ftp).

Como Archie provee un índice de los servidores a contactar, las búsquedas no están restringidas por la naturaleza jerárquica de los nombres de los host de Internet.

b). La otra base de datos, whatis, contiene los nombres y descripciones de paquetes de software, documentación y otra información disponible en Internet. Las entradas en esta base son cadenas de caracteres conteniendo claves y una descripción asociada.[DAN]

Los usuarios realizan la búsqueda y esta es aplicada tanto a las claves como a la descripción.[DAN]

Actualmente es mantenida manualmente. La información es obtenida de diferentes fuentes como por ejemplo de publicaciones de artículos en USENET, o mediante mensajes de correo electrónico generados por el autor.[DAN]

II. Arquitectura Archie[DAN]

La base de datos con los nombres de los archivos es mantenida por:

- DGC (data gathering component): componente de recolección de datos. Se basa en los administradores de servidores FTP para conocer nuevos archivos disponibles mediante ftp.**
- DMC (data maintenance component): componente de mantenimiento de datos.**

El cliente Archie accede a ambas bases a través de la UAC (user access component) componente de acceso de usuarios.

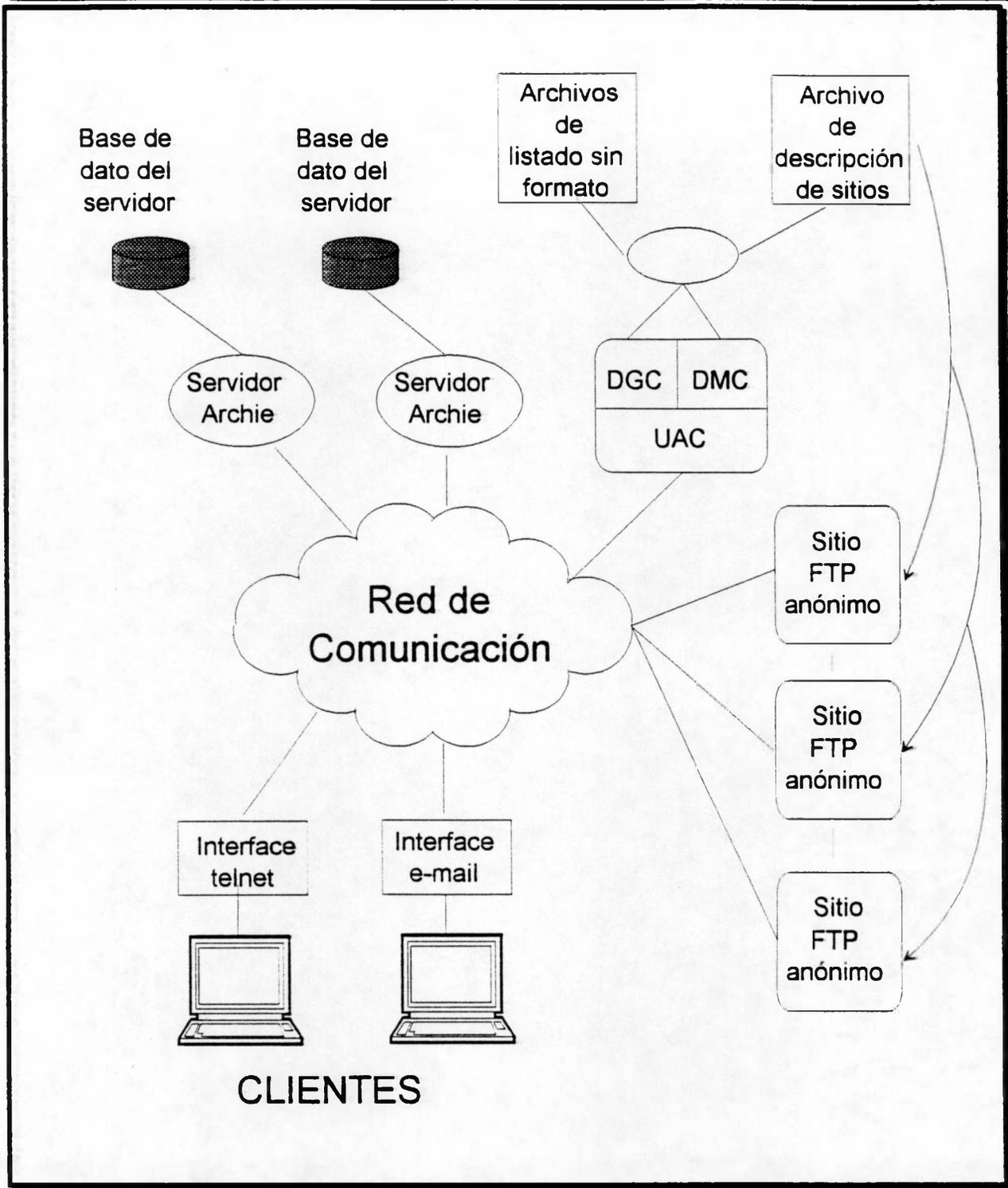


Fig B : Arquitectura Archie [Dan]

Cada vez que un nuevo sitio ftp es reportado, se agrega una nueva entrada en el archivo de descripción de sitios. Este lista los sitios ftp conocidos. Periódicamente, el DGC se conecta con cada sitio FTP conocido, y realiza un listado recursivo del contenido disponible en el directorio. Esta información es mantenida en un servidor Archie, en archivos de listado sin formatos y es procesado por el DMC convirtiendo esos datos en bruto en formateados para que puedan ser agregados a la base de datos con los nombres de archivos.

III. Como acceder a Archie [KOC/93. Gil/94]

El acceso puede hacerse en forma directa o a traves de otras herramientas como Gopher o WWW. Existen tres maneras para el ingreso directo: vía un cliente, vía el comando telnet o vía correo electrónico. El ingreso mediante otras herramientas lo veremos posteriormente.

1) Vía Telnet.

Primero debemos seleccionar el servidor Archie. En forma de ejemplo podemos ver mas adelante una lista de servidores Archie en el mundo. Debido al cambio constante, estos servidores pueden diferir.

archie.au*	139.130.4.6	Australia
archie.edvz.uni-linz.ac.at*	140.78.3.8	Austria
archie.univie.ac.at*	131.130.1.23	Austria
archie.uqam.ca*	132.208.250.10	Canada
archie.funet.fi	128.214.6.100	Finland
archie.th-darmstadt.de*	130.83.22.60	Germany
archie.ac.il*	132.68.6.15	Israel
archie.unipi.it*	131.114.21.10	Italy
archie.wide.ad.jp	133.4.3.6	Japan
archie.kr*	128.134.1.1	Korea
archie.sogang.ac.kr*	163.239.1.11	Korea
archie.rediris.es*	130.206.1.2	Spain
archie.luth.se*	130.240.18.4	Sweden
archie.switch.ch*	130.59.1.40	Switzerland
archie.ncu.edu.tw*	140.115.19.24	Taiwan
archie.doc.ic.ac.uk*	146.169.11.3	United Kingdom
archie.unl.edu	129.93.1.14	USA (NE)
archie.internic.net*	198.48.45.10	USA (NJ)
archie.rutgers.edu*	128.6.18.15	USA (NJ)
archie.ans.net	147.225.1.10	USA (NY)
archie.sura.net*	128.167.254.179	USA (MD)

Sites marked with an asterisk "*" run archie version 3.0

Es aconsejable realizar la selección considerando la distancia, la congestión de la red a fin de evitar sobrecarga en esta. Ningun servidor es de proposito especial o es mejor, por lo tanto no hay necesidad de aumentar la congestión.

Luego de la selección, el comando <telnet nombre-archivo> nos permite transportarnos al servidor. El nombre del usuario es <archie>. Si el servidor esta sobrecargado no permitirá el ingreso de nuevos usuarios por lo que nos enviará un mensaje alertandonos de esta situación y cortará la conexión. Pero si tenemos suerte nos dejará ingresar y trabajar con él.

Archie fue diseñado para responder alguna de las siguientes preguntas[KOC/93]:

- *¿ Existe un archivo o directorio con un cierto conjunto de caracteres en su nombre? => Búsqueda por nombre*
- *¿ Que archivo se refiere al tema X ? => Búsqueda por descripción de archivo.*
- *¿ Que host FTP está actualmente monitoreado por Archie? ¿Donde están localizados? ¿Que archivos en particular contiene el host? => Búsqueda por sitio.*

La búsqueda mas simple es buscar una cadena de caracteres. Para ello podemos usar el comando <find cadena>. Archie buscará en su base de datos todos los archivos que concuerden con la cadena dada.

Esta concordancia varia según hallamos especificado anteriormente. Por default asume el valor "sub". Veamos en la siguiente tabla algunos tipos de concordancia.

<i>exact</i>	<i>exactamente</i>	<i>El nombre del archivo debe concordar exactamente con el modelo.</i>
<i>sub</i>	<i>subcadena</i>	<i>El modelo debe concordar con el nombre completo o con parte de este. No importa si esta en mayúscula o minúscula.</i>
<i>subcase</i>	<i>subcadena</i>	<i>Idem anterior pero debe coincidir las mayúsculas y minúsculas.</i>
<i>Regex</i>		<i>Permite utilizar los caracteres "*" , "?" , etc.</i>

Originalmente el comando de búsqueda se llamó <prog>. Realmente describía que era lo que estaba buscando: programas. Con el paso del tiempo Archie ofrecía más que programas, ofrecía documentos, directorios y otro tipo de información. Para ese entonces, la palabra prog era difícil de relacionar y por lo tanto de recordar. Se vio que lo mas conveniente era utilizar la palabra <find> (búsqueda). En un ambiente de comandos es la primera palabra que nos viene intuitivamente a la mente. Por lo que algunos sitios Archie renombraron el comando <prog> por <find>. Aunque los usuarios que estan acostumbrados a utilizar el viejo nombre pueden hacerlo.

Ahora veamos la búsqueda por descripción de archivo. Para ello debemos utilizar la base de descripción de software (Software Description Database) y el comando para realizar la búsqueda es <whatis cadena>. La información que nos proporciona es el nombre de archivo, datos adicionales y una descripción breve para darnos una idea de que se trata el archivo. No nos da la ubicación de este y para saber eso debemos utilizar el comando find.

Para realizar una búsqueda con gran posibilidad de éxito, debemos tener una idea de como se llama el archivo que buscamos o tener suerte de que el nombre sea obvio. Pero si lamentablemente no estamos en ninguna de los dos casos anteriores, lo mas conveniente es usar el comando whatis para que nos oriente en nuestra travesía. Parece que este comando fuera mágico pero lamentablemente no lo es y veamos porque. La base de descripción de software padece de varios inconvenientes. El principal es que es dependiente de la descripción que haga el creador del archivo. Cada persona puede describir una cosa de diferente manera y a veces estas descripciones no tienen nada que ver con el archivo. Otro problema es que la base de descripciones esta a parte de la base principal donde Archie realiza su búsqueda, por lo que pueden existir diferencias entre ellas. Finalmente, la Base de datos tiene problemas de actualización debido a que archivos pueden haber sido removidos y este cambio no se refleja inmediatamente o el usuario que le dé de alta a un archivo FTP no envíe la descripción de este.

El último tipo de búsqueda es por sitio[KOC/93]. Aquí el objetivo es saber que sitios FTP son monitoreados por Archie. En estas circunstancias tenemos dos posibles vistas. Una es querer saber todos los sitios y la otra es pararnos en uno en particular e investigar que archivos tiene. Los comandos que debemos utilizar son en el primer caso <list> y en el segundo es <site>. Debemos tener en cuenta que un sitio puede tener miles de archivos disponibles, por lo tanto es aconsejable acompañar al comando site con alguna cadena bastante descriptiva que pueda filtrar el resultado de la búsqueda.

ii) Vía un cliente .

Ingresar en el mundo Archie mediante el comando telnet es una buena idea pero tal vez no es la más adecuada. Veamos porque. Archie cuenta con una limitación en el número de usuarios que pueden trabajar simultáneamente. Esa cantidad esta dada para evitar una sobrecarga en los servidores especialmente en las horas picos. Cada conexión requiere de recursos de computación los que a mayor cantidad de usuarios la performance del sistema decae.

Esta restricción no es impuesta sobre el cliente. Por lo que podemos evitarnos de esta manera la pérdida de tiempo ocasionada por los rechazos de los ingresos vía telnet. Además es mucho mas rápido y ayuda a decrementar la sobrecarga en la red.

Clientes Archie son disponibles para muchos sistemas operativos como por ejemplo Unix, PC, VMS entre otros. Tanto los servidores como los clientes estan disponibles gratuitamente vía FTP.

Existe una restricción al usar el cliente, no podemos acceder de esta forma a la base de datos whatis por lo que se deberá utilizar otras vías como telnet o correo electrónico.

iii) Archie vía correo electrónico.

Existe una tercera forma de acceder a Archie y es a través del correo electrónico. Esta ruta es indirecta y mucho mas lenta debido a los retardos propios del correo electrónico y al procesamiento del mensaje. Pero el usuario tiene la ventaja de no perder tiempo mientras espera el resultado de la búsqueda. Además es una excelente opción para aquellos usuarios que no poseen de un cliente ni de una conexión telnet.

Se puede escribir un mensaje a cualquiera de los servidores Archie. La dirección se forma por el nombre de usuario archie mas la dirección de la maquina deseada. No es necesario ponerle subject al mensaje ya que el comando debe ir en el cuerpo del mensaje

Para obtener el archivo luego de encontrar su posición a través de Archie, debemos utilizar el comando FTP para obtener una copia en nuestra máquina.

Si mandamos un mensaje como el siguiente,

dirección: archie@archie.au
Subject: <no es necesario>
Cuerpo
prog archie

Archie nos enviara una respuesta como lo muestra en la pagina siguiente.

La respuesta de Archie fue:

```

From Archie.AU!archie-errors Tue Apr 5 19:28:06 1994
Received: from plaza.sarnet.EDU.AU ([129.130.4.6]) by secyt.gov.ar with SMTP id <18105>; Tue, 5 Apr 1994 19:37:52 -
0400
Received: (archie@localhost) by plaza.sarnet.EDU.AU (8.6.8/8.6.5) id GAA15160; Wed, 6 Apr 1994 06:08:56 +1000
To: Cecilia Sarmiento <ceci@secyt.gov.ar>
From: (Archie Server) archie-errors@Archie.AU
Reply-To: (Archie Server) archie-errors@Archie.AU
Date: Tue, 5 Apr 1994 16:08:00 -0400
Subject: archie [prog archie] part 1 of 1
Status: RO

>> path Cecilia Sarmiento <ceci@secyt.gov.ar>

>> prog archie
# Search type: exact.
Host nic.switch.ch (120.59.1.40)
Last updated 17:15 5 Apr 1994

Location: /file_server
DIRECTORY drwxrwxr-x 512 bytes 16:37 3 Dec 1993 archie

Location: /mirror/vms/spc/multinet/contributed-software/applications/archie
DIRECTORY drwxrwxr-x 1024 bytes 12:40 6 Dec 1993 archie

Host yuma.acna.colostate.edu (129.82.100.64)
Last updated 08:37 5 Apr 1994

Location: /software.lbmpe
DIRECTORY drwxr-xr-x 512 bytes 00:00 20 Feb 1992 archie

Host wncc-fuk.wide.ad.jp (133.4.20.2)
Last updated 08:22 5 Apr 1994

Location: /pub/misc
DIRECTORY drwxrwxr-x 512 bytes 00:00 26 Nov 1991 archie

DIRECTORY drwxrwxr-x 1024 bytes 19:51 12 Mar 1994 archie

Host softu1.ncu.edu.tw (140.115.19.11)
Last updated 00:13 5 Apr 1994

Location: /pub/386BSD/0.1-ports/utlis
DIRECTORY drwxr-xr-x 512 bytes 00:00 5 May 1992 archie

Host nuscc.nus.sg (192.169.33.2)
Last updated 23:21 4 Apr 1994

Location: /pub/ftp-llst
FILE -rw-r--r- 12899 bytes 23:00 6 Jun 1991 archie
>> quit

```



BIBLIOTECA
FAC. DE INFORMÁTICA
U.N.L.P.

CAPITULO III

GOPHER ~ VERONICA ~ JUGHEAD

Gopher es una herramienta que permite recorrer el mundo de la información en Internet a través de un conjunto de menús de recursos que estan disponibles en la red de servidores de Gopher. La base es buscar el recurso de interés, navegando a través de distintas opciones representadas por menús y seleccionando sin importar en que lugar del mundo está, ni que formato tiene.

Aquí la información esta organizada jerárquicamente mediante menús que dan la orientación de que camino seguir en la búsqueda del recurso. De esta forma ayudar al usuario ya que no tiene que recordar comandos, solo selecciona que camino seguir. No debemos olvidarnos que la estructura jerarquica es la mas natural para la mayoría de los usuarios debido a que este es uno de los primeros conceptos que tiene que entender al utilizar directorios por lo que es muy fácil utilizar este sistema.

Esta herramienta fue creada en el año 1991 en la Universidad de Minnesota y surge a la necesidad que plantea su campus. La Universidad contaba con una gran diversidad de computadoras (workstations, PC, mainframe, etc.) las cuales requerian un conocimiento un poco elevado para manejarlas y mas aún, obtener la información que se necesitaban la cual estaba dispersa por todas ellas. La gran mayoría de los usuarios no contaban con tal destreza por lo que Mark McCahill y su grupo de programadores desarrollaron esta herramienta con el proposito de ayudar a su campus. Rapidamente se hace muy popular y lo que era una ayuda para un campus se transformó en una ayuda para los usuarios de Internet. [Gil/94]

El nombre dado fue en honor a la mascota de la Universidad que es el Golden Gopher una especie de ardilla. Pero Gopher tiene varios significados que son muy apropiados para esta herramienta. Como por ejemplo, este animalito recorre túneles subterráneos hasta encontrar la planicie. La analogía con el

Gopher Internet es que este puede atravesar las rutas Internet hasta conseguir la información requerida[KOC/93]. Además "Go-fer" es un americanismo que significa "alguien que busca cosas o provee servicios a otras personas".[KOC/93]

Cuenta también con la ayuda de otros servicios como Archie para hallar sitios FTP, News, Wais (el cual veremos en el próximo capítulo), X500(ver glosario), y directorio de páginas blancas para la búsqueda de usuarios (ver glosario).

I. Modelo de los datos

El mundo Gopher consiste en un conjunto de servidores que contienen recursos representados a través de menús. Nos permite ver a la información disponible como una estructura de directorio infinita disponible a través de menús como si estuviera todo almacenado en nuestra máquina[Koc/93].

La utilización del modelo de sistema de archivos para la localización de recursos es adecuado por varias razones. La primera, como habíamos dicho anteriormente, es una estructura muy familiar para la mayoría de usuarios. Segundo, puede expresarse fácilmente por una sintaxis simple. Tercera razón, este sistema fue desarrollado en una universidad, en donde mayormente la información esta compuesta por documentos de texto almacenados en directorios en computadoras personales por lo que el proceso de publicación al Gopher es directo. Y por último, es extensible, esto quiere decir que se puede ir agregando archivos en la estructura muy fácilmente. [Alb/92]

Los servidores ofrecen la siguiente información:

• Archivos de Texto: Estos están indicados en el menú con un punto final. Para leerlos solo se necesita seleccionarlos. Existen 2 formas para poder capturarlos, enviarlo por correo electrónico o guardarlo en un archivo: Gopher graba el archivo en el directorio donde el cliente Gopher empezó a correr.

• Servidores de páginas blancas: los servicios de directorio de páginas blancas ("white pages") son esencialmente guías telefónicas electrónicas. Gopher tiene la habilidad de acceder a este servicio.

• **Otros servidores:** el menú de Gopher contiene un entrada para poder movernos a otro servidor. [KRO/94] **Búsquedas indexadas** :Cuando se selecciona una búsqueda indexada se podrá ingresar las palabras que se utilizarán como claves. Gopher enviará un requerimiento al servidor donde el material indexado está almacenado y luego mostrará archivos conteniendo la palabra clave en sus títulos o como contenido. [KRO/94]

Quando el usuario selecciona un directorio corriente ve un menú de todo lo que se encuentra en este. Quando realizamos búsquedas indexadas el trabajo de indagar es similar excepto que el menú solo muestra un subconjunto del contenido del directorio. El subconjunto que se ve está basado sobre una búsqueda con claves. Quando el usuario da la orden de buscar una clave, Gopher busca y construye un menú cliente que solo contiene los ítems que concordaron con el criterio de búsqueda que propuso el usuario. [KRO/94]

Existen dos problemas. Uno es que la interface de Gopher es muy general por lo que trae algunas confusiones. El usuario no tiene idea en que computadora o software está realmente haciendo la búsqueda. Gopher puede estar haciendo la búsqueda a través de servidor Archie, servidores Wais y en otros. Cada uno de estos servidores tiene sus propias reglas de búsqueda e interpreta la clave en forma diferente. El segundo, Gopher tiende a reducir las habilidades de búsqueda de diferentes servicios a la intersección de sus características. Por ejemplo se puede acceder a Wais a través de Gopher. Las búsquedas de Wais son extremadamente poderosas, mucho de ese poder se pierde pues no se pueden utilizar todas las características a través de Gopher. [KRO/94]

• **Telnet a través de Gopher:** Gopher puede conectar al usuario con el recurso utilizando telnet como una interface. Quando se selecciona este tipo de recurso hay que tener cuidado pues Gopher pierde el control de la sesión una vez que se dispara el telnet. [KRO/94]

Quando el usuario termina en forma normal (log out) el telnet, Gopher volverá al menú donde el usuario eligió este recurso. [KRO/94]

• ***FTP a través de Gopher: Se puede utilizar Gopher como alternativa a una interface ftp. Permite mover archivos de servidor de ftp anónimo a una computadora corriendo un cliente Gopher. [KRO/94]***

Con solo seleccionar en el menú el ítem específico de ftp, el usuario puede navegar a través de series de menús para encontrar el servidor y el archivo deseado. [KRO/94]

Algunos servidores Gopher utilizan al sistema Archie para hallar los servidores FTP anónimos. Estos construyen un recurso indexado, accesible por menú, de los servidores ftp anónimos de todo el mundo. En realidad Gopher esconde los detalles de Archie, el usuario no ve que está accediendo a otro tipo de servicio. Gopher puede realizar dos tipos de búsquedas Archie: concordancia exacta de la cadena de caracteres pedida y concordancia de una subcadena. [KRO/94]

Gopher trabaja bien para archivos de texto pero con archivos binarios tiene algunos problemas. Para manejar estos archivos apropiadamente, Gopher debe suponer que tipo de archivo es basado en una extensión del nombre (por ejemplo.tar.Z). Cada cliente lo maneja de distinta forma. Por ejemplo, el cliente Macintosh trabaja perfectamente con archivos binarios y lo descomprime si ellos terminan con una extensión ".bqx". La versión de Unix lo traslada bien pero la versión X no puede transferir estos tipos de archivos binarios. [KRO/94]

Si Gopher no lo traslada bien o directamente no lo hace, este puede informar al usuario cual es el host que contiene al archivo buscado y además indicarle cual es la ruta para llegar a él. Entonces, el usuario puede utilizar un acceso FTP anónimo, llegar al directorio deseado y una vez ahí utilizar el comando FTP para trasladar el archivo a su sistema. [KRO/94]

En forma reducida:

<i>Tipo de recurso</i>	<i>Ejemplos Específicos</i>
<i>Textos y archivos</i>	<i>Libros electrónicos</i> <i>Archivos de USENET news group, listas</i> <i>Archivos de sitios FTP</i>
<i>Recursos Telnet</i>	<i>Catálogos de bibliotecas on-line</i> <i>Directorios de números de teléfonos</i> <i>Directorios de direcciones de correo electrónico</i> <i>Sistemas de Información de Campus (CWIS)</i>
<i>Búsquedas en Bases de Datos</i>	<i>Búsquedas de textos completo de archivos almacenados en Gopher</i> <i>Búsquedas en sitiosarchie</i> <i>Wais</i>
<i>Misceláneas</i>	<i>Gráficos, sonido, imágenes</i>

Todos los servidores estan conectados unos a otros proporcionando una continuidad de la información alrededor del mundo.

III. Interface Gopher

Los menús constan de dos partes principales. Una es una indicación del directorio actual en donde el usuario está parado. La otra es una serie de opciones de recursos que el usuario puede elegir con solo seleccionarlo. La forma de presentar a esto distintos tipos de recursos es mediante una marca especial al final de la línea en el menú

```

Internet Gopher Information Client v1.12s
News

1. Daily Texan (University of Texas, Austin)/
2. Minnesota Daily/
3. National Weather Service Forecast/
----> 4. Reacent Earthquakes.
5. SOUND, News and Arts Newspaper//
6. Technolog (Institute of technology, U. of Minn)/
7. Upi News/
8. USENET News (from Michigan States) /

Press ? for Help, q to quit, u to go up
Page:1/1
    
```

Tipo de Recurso	Marca al final de la línea
<i>directorio</i>	/
<i>archivo de texto</i>	.
<i>archivo binario</i>	<BIN>
<i>servidores de white pages</i>	<CSO>

<i>directorio indexado</i>	<?>
<i>ftp</i>	
<i>telnet</i>	<TEL>
<i>sonido digitalizado</i>	<

Tabla de ejemplo

IV. Arquitectura Gopher. [DAN]

Está compuesta de clientes y servidores comunicándose a través del protocolo Gopher que esta implementado sobre TCP/IP. (Ver figura A.)

La raíz de la jerarquía Gopher, está almacenada en el host de la Universidad de Minnesota. Este es el directorio default recuperado por el cliente cada vez que es invocado, aunque este puede ser configurado para que se conecte a otro host. El raiz conoce todos los servicios del nivel superior. En esta arquitectura existe un único servidor de nivel superior por cada organización como por ejemplo un campus de una universidad, una dependencia de gobierno o una institución. Los servidores de nivel inferior pueden ser enlazados al correspondiente de nivel superior. De esta manera, una vez que el usuario encuentra a un servidor de alto nivel puede navegar a través de esta jerarquía siguiendo los lazos a los servidores de nivel inferior.

Existen actualmente dos protocolos utilizados por los servidores. El protocolo Gopher, es el primero que salio al mercado y el nuevo llamado Gopher+.

Veamos primero el protocolo Gopher y luego una breve descripción del Gopher+.

Protocolo Gopher

I El modelo internet del protocolo Gopher.[ALB/92]El cliente se conecta con un servidor y le envía a este un selector (una línea de texto que puede ser vacía) vía una conexión TCP en un port conocido. El servidor responde con un bloque de texto terminando con una línea que solo contiene un punto y cierra la conexión.

Veamos un simple ejemplo (Ver tabla próxima página) de una interacción cliente/servidor. Asumamos que el servidor Gopher sabe bien en que port debe escuchar. La única configuración que el cliente retiene es su nombre de servidor y el numero de port (en el ejemplo siguiente el nombre del servidor: rawBits.micro.umn.edu ; número de port: 70). El # significa tabulador.

Veamos los campos de cada línea

Primera sección comprendida hasta el primer tabulador lo podemos dividir en 2 partes:

- 1. el primer caracter de la línea describe un documento, servidor CSO, directorio o error:*

Tipo	Caracter
documento	0
directorio	1
servidor CSO	2
error	3
...etc...

<p>Cliente:</p>	<p>Abre la conexión con rawBits.micro.umn.edu en el port 70</p>
<p>Servidor:</p>	<p>Acepta conexión pero no envía nada</p>
<p>Cliente: <CR> <LF></p>	<p>Envía una línea vacía que significa "lista lo que tienes"</p>
<p>Servidor:</p> <p>0About internet Gopher#Stuff: About us # rawBits.micro.umn.edu#70</p> <p>1Around the University of Minnesota#Z,5692,Aum#underlog.micro.umn.edu#70</p> <p>1Microcomputer News & Prices#Prices/#pserver.bookstore.umn.edu#70</p> <p>1Courses, Schedulers, Calendar##events.ais.umn.edu#70</p> <p>1Students-Staff Directories##uinfo.ais.umn.edu#70</p> <p>1Departmental Publications#Stuff:DP:#rawBits.micro.umn.edu#70</p> <p>.....etc...</p> <p>.</p>	<p>Envía una serie de líneas cada una terminado con un CR LF</p> <p>Pone un punto</p> <p>Cierra la conexión.</p>

Cuarta sección se halla el número de port donde conectar.

Veamos el ejemplo:

Cuarta sección: 70

El cliente mostrará en pantalla lo siguiente:

- 1. About internet Gopher.**
- 2. Around the University of Minnesota - Office and Services/**
- 3. Courses, Schedules, Calendars/**
- 4. Events/**
- 5. Microcomputer News & Prices/**
- 6. Student-Staff Directories/**
- 7. University Relations Informations and Forms/**
- 8. Weather for the Twins Cities.**

Para el usuario es transparente en que máquina y donde reside el recurso.

Supongamos que el usuario selecciona la línea "Microcomputers News & Prices".

Veamos la interacción entre el cliente y servidor:

<p>Cliente:</p>	<p>Abre la conexión con</p> <p><i>pserver.bookstore.umn.edu</i></p> <p>u</p> <p>en el port 70</p>
<p>Servidor:</p>	<p>Acepta conexión pero no dice nada</p>
<p>Cliente: Prices/</p>	<p>Envía la línea para ubicar el directorio</p>
<p>Servidor:</p> <p><i>0About Prices#Prices/About us #</i> <i>pserver.bookstore.umn.edu#70</i></p> <p><i>0Macintosh Prices#Prices/Mac#pserver.bookstore.umn.edu#70</i></p> <p><i>0ZEOS Prices#Prices/ZEOS#pserver.bookstore.umn.edu#70</i></p> <p><i>0IBM Prices#Prices/Ick#pserver.bookstore.umn.edu#70</i></p> <p><i>0Printer & Peripheral</i> <i>Prices#Prices/PPP#pserver.bookstore.umn.edu#70</i></p> <p><i>.....etc...</i></p> <p>.</p>	<p>Envía una serie de líneas cada una terminado con un CRLF.</p> <p>Pone un punto.</p> <p>Cierra la conexión.</p>

iv. El cliente:[ALB/92]

Un cliente simplemente envía una cadena de caracteres a un servidor si desea obtener un documento o ver el contenido de un directorio. Puede suceder que un host tenga punteros a otros hosts, resultando el sistema de archivos en una estructura tipo grafo (más que en una estructura arbolada) ya que estos punteros pueden apuntar hacia atrás, al nivel superior.

El cliente guardará en una pila la ubicación que este ha visitado en la búsqueda de un documento. El usuario puede volver hacia atrás de la ubicación actual con solo desapilar el stack.

Un servicio, en particular aquellos que son críticos, puede estar duplicado más que estar en un único servidor. Si un cliente no puede comunicarse con un servidor podría tratar de comunicarse con un duplicado. Lo ideal sería que tomara uno al azar para esparcir la carga a través de los servidores.

v. El servidor[ALB/92]

La cadena de caracteres que es enviada al servidor para recuperar un documento (u otro servicio) puede ser una ruta a un archivo o a un directorio. Este puede ser el nombre de un script, una aplicación o una consulta que produce el retorno de un documento o directorio. El servidor descarta los CR-LF y los tabuladores. Siguiendo a un tabulador opcional esta un descriptor de día-hora (AAAAMMDDhhmmss)

El tabulador y el descriptor existen para una eficiente búsqueda por texto-completo (full-text). Si el tabulador y el descriptor están presentes, el servidor retornaría solo los ítems que han sido modificados desde el día y la hora que especifica el descriptor. Si el servidor no puede implementar el filtrado, esto solo lo descarta. Veamos un ejemplo:

Osiris <CR> <LF>

Envía :

- **si Osiris es un selector de directorio => envía un listado de directorio.**
- **si Osiris es un selector de archivo => envía un archivo.**

Osiris <TAB>19910315000000<CR> <LF>

Retorna :

- **si Osiris es un selector de directorio => envía un listado todos los subdirectorios pero incluye solo los archivos que han sido modificados desde el 15 de marzo de 1991.**
- **si Osiris es un selector de archivo => envía el archivo solo si ha sido modificado desde el 15 de marzo de 1991.**

Ademas de los servidores normales de Gopher, existen dos tipos de servidores especiales. Uno es Servidores CSO. Una entrada en el listado de directorio puede apuntar a un servidor CSO. Si este ítems es seleccionado, el cliente utilizara el protocolo CSO cuando este se conecta al host apropiado.

Veamos rápidamente que es un servidor CSO. Es un sistema de directorio que provee una forma de localizar usuarios conectados en Internet. Por su similitud con las guías telefónicas que usa páginas blancas para el listado de usuarios y páginas amarillas para el comercial, la comunidad técnica lo llamo "servicio de páginas blancas". Debido a varias razones no hay un único servicio de directorio en todo Internet, existen diferentes tipos entre los que hallamos a CSO. Un gran número de colegios y universidades usan este servicio para poner en línea la información de alumnos y personal. Es necesario para usarlo un cliente llamado ph pero muchos acceden a este directorio mediante Gopher. [KRO/94]

Los otros son Servidores de búsqueda de texto completo. Un servidor de búsqueda de texto completo es un servidor de propósito especial que conoce el esquema que utiliza Gopher para recuperar archivos. Estos mantienen un índice de texto completo del contenido de los documentos de los servidores Gopher en algún dominio específico. Estos fueron implementados usando NeXTstations para tomar ventaja del software de búsqueda contenidos en este tipo de computadora. También están disponibles servidores de búsquedas para Unix basados en el tipo de búsqueda que utiliza Wais.

Para usar varios servidores de índice (en vez de utilizar uno solo) Gopher está capacitado para construir y buscar índices en paralelo. La tarea puede ser dividida en piezas pequeñas (por ejemplo, actualizar solo una porción de los índices o búsquedas de índices parciales en paralelo). Esparciendo esta tarea sobre varias (pequeñas, baratas y rápidas) workstations, se puede tomar ventaja de un paralelismo refinado. El cliente no tiene que darse cuenta de esto. Este, solo envía una cadena de caracteres para buscar y recibe una lista de documentos que contienen las palabras buscadas.

Para implementar una búsqueda algunas máquinas pueden mantener índices de texto completo sobre el contenido del texto de los documentos. Este tipo de servidor responde al requerimiento del cliente con una lista de los documentos que contienen una o más de las palabras. El cliente envía al servidor la cadena de caracteres selectora, un tabulador, las palabras a buscar. El servidor retorna el equivalente a un listado de directorio para los documentos que concuerdan con el criterio de búsqueda.

Las palabras "and", "or" y "not" son reservadas como operadores lógicos y las expresiones con estos operadores, son evaluadas de izquierda a derecha.

vi. Conjunto de tipo de ítems[ALB/92]

El cliente decide que ítems están disponibles solo mirando el primer caracter de cada línea del listado enviado por un servidor. Aumentando esta lista podemos extender el protocolo.

Caracter	Descripción del Ítem
0	Archivo
1	Directorio
2	Servidor CSO de guía telefónica
3	Error
4	Archivo Binhexa Macintosh (?)
5	Archivo Binario DOS
6	Archivo Codificado(unicode) Unix
7	Servidor de búsqueda indexada
8	Sesión telnet
9	archivo binario
+	Servidor redundante (tiene la misma información anterior)

Protocolo Gopher+ [Mor/94]

Desde el protocolo Gopher original, muchas cosas fueron cambiando como por ejemplo correr programas remotos y mostrar gráficos. Algunos administradores Gopher han modificado el protocolo original para cubrir sus propias necesidades.

La Universidad de Minnesota introdujo un nuevo protocolo llamado Gopher+. Este es compatible con la versión anterior e incluye las funciones de: agregar datos a los servidores, permite al administrador del servidor generar formularios para consultar a los usuarios y actuar sobre las respuestas, enviar mensajes de correo electrónico.

VI. Servidor Gopher: [GOP/92]

La forma de poner en servicio un servidor Gopher es muy fácil, solo hace falta compilar el servidor. Luego crear un directorio propio para el sistema en donde se almacenaran los datos que van a ser accedidos por los clientes. Por último correr el demonio de Gopher. Un demonio es un programa que permanece dormido esperando que un cliente requiera algún dato del servidor en cuyo momento se despierta, realiza su tarea según sea lo que necesite el usuario. Cuando termina, envía los datos al cliente y vuelve a dormir hasta el próximo requerimiento.

También es fácil agregar datos nuevos. Debemos ver si la información puede ser añadida en algún directorio creado anteriormente y actualizar los archivos que Gopher utiliza para su tarea. Si no es posible, se debe crear un nuevo directorio, crear los archivos que necesita Gopher y agregar la información.

Para mantener la continuidad de los datos, se debe enlazar los servidores. Para lograr esto. Gopher necesita un archivo de uso especial llamado link el cual contiene la dirección IP del otro servidor a contactar en el caso que el usuario lo requiera.

IV. Como acceder a Gopher.[KOC/93]

Gopher puede ser accedido mediante dos formas directas. Una es vía un cliente y la otra es utilizando el comando telnet. También podemos ingresar a través de WWW.

i. Gopher vía un software cliente

Existen clientes Gopher de libre distribución para muchos tipos de sistemas operativos, como por ejemplo: para sistema Operativo UNIX existe el cliente Gopher. Para más sofisticados hardware como los que utilizan conexiones directas Internet, existe el Xgopher para los sistemas X windows. Para Macintosh tenemos el cliente TurboGopher que puede utilizar el protocolo Gopher+. También encontramos clientes para Dos, Window, Next, VMS. El

software puede ser encontrado y adquirido en forma gratuita via FTP en la dirección boombox.micro.umn.edu. [Gil/94]

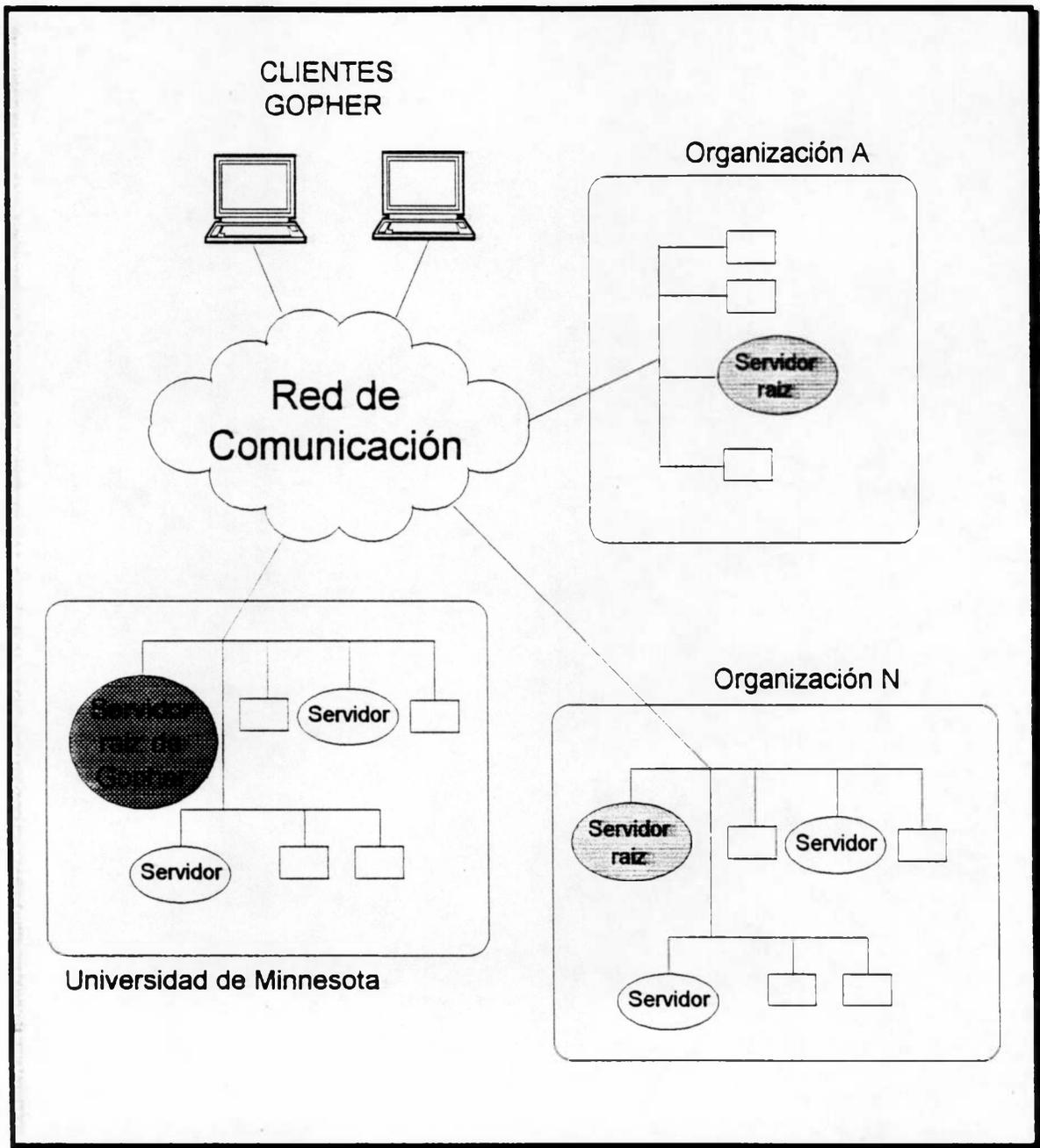


fig A: Arquitectura Gopher.

ii. Gopher vía telnet

Si no está instalado el cliente en la computadora se puede realizar un telnet a una que ofrezca este servicio. Como por ejemplo:

<i>Nombre del Host</i>	<i>Login</i>	<i>Ubicación</i>
<i>consultant.micro.umn.edu</i>	<i>gopher</i>	<i>Minnesota, EE.UU.</i>
<i>gopher.uiuc.edu</i>	<i>gopher</i>	<i>Illinois, EE.UU.</i>
<i>gopher.uwp.edu</i>	<i>gopher</i>	<i>Wisconsin, EE.UU.</i>
<i>info.anu.edu.au</i>	<i>info</i>	<i>Australia</i>
<i>gdunix.gd.chalmers.se</i>	<i>gopher</i>	<i>Suiza</i>
<i>ecnet.ec</i>	<i>gopher</i>	<i>Ecuador</i>
<i>gopher.secyt.gov.ar</i>	<i>gopher</i>	<i>Argentina</i>

Listas de algunos servidores Gophers

VIII. Veronica. .

La cantidad de servidores Gopher creció tan abruptamente y con el aumento la dificultad de hallar la información necesitada. Fue necesario crear una herramienta para permitir la búsqueda de información dentro del mundo Gopher. Con el propósito de resolver este problema se crea veronica. [Mor/94]

La primera versión fue desarrollada en Noviembre de 1992 por Steve Foster y Fred Barrie de la Universidad de Nevada, Reno.[HAH/94]. El nombre es una sigla que significa "Very Easy Rodent-Oriented Net-wide Index to Computerized Archived". Aunque realmente es un juego de palabras ya que Veronica es el nombre de la novia de Archie en la tira comica que hablamos en el capitulo anterior. [Mor/94/

Veronica realiza la búsqueda en un índice que actualiza semanal o quincenalmente de todos los ítems en los menús de lo servidores Gopher que estan reconocidos por el servidor localizado en la Universidad de Minnesota. La forma de trabajar es basicamente la misma que la de Archie, busca en una base de datos propia. [Gil/94]

La forma de acceder a este servicio es a traves de Gopher y solo se debe seleccionar del menú de directorio principal la opción "...using Veronica". Luego de seleccionarlo se verá uno o más ítems sobre búsquedas en la base de datos. Simplemente se selecciona un índice de búsqueda y se especifica la palabra clave. [HAH/94]

La diferencia con la búsqueda mencionadas en párrafos anteriores es que Veronica busca en todas partes para hallar el ítem, en cambio la búsqueda indexada solo la realiza en la base de datos seleccionada por el usuario.[HAH/94]

Unos de los valores de Veronica es que permite una vista del espacio de Gopher según claves. HAH/94]

En resumen, este sistema soporta búsqueda solo en los nombres de los directorios ó en los títulos.

LX. Jughead[Gil/94][

Es una herramienta similar a Veronica pero la búsqueda se restringe a una parte del espacio de Gopher seleccionada por el usuario.

Fue desarrollada por Rhett Jones (Jonzy) en el centro de computos de la Universidad de UTAH y su primer versión fue en marzo de 1993. Nuevamente su nombre es un juego de palabras debido a que es una sigla que significa "Jonzy's

Universal Gopher Hierarchy Excavation And Display” pero tambien, como el caso de veronica, Jughead es otro de los personajes amigo de Archie.

El rol de un servidor Jughead es mantener una base de datos con todos los ítems de los menús dentro de ese sitio Gopher y permitiendo buscar en el

Veamos un ejemplo: un usuario está interesado en libros y bibliotecas de una universidad en particular. Podemos utilizar Veronica y generar una búsqueda. Sin embargo existen varios problemas utilizando esta herramienta: puede encontrar demasiados ítems con la descripción pedida; Veronica tiene un límite de construcción de los ítems que encuentra sobre un tema en particular, por lo tanto podría exceder su límite antes de llegar a la universidad deseada y el tiempo que le tomara a Veronica puede ser grande. Pero si el usuario usara Jughead, este podría restringir la búsqueda a la universidad deseada y acelerarla. La única limitación es que el servidor Jughead ya tendrá que tener especificada ese espacio del Gopher.

La forma de acceder a este servicio es parecido a Veronica: seleccionando del menú de Gopher el ítem apropiado.

La búsqueda es similar a las realizadas por veronica, se ingresa una o más claves y Jughead las utiliza para la búsqueda. El resultado se envía en forma de menú, por lo tanto el usuario puede seleccionarlo de manera regular.

CAPITULO IV

WAIS

Wais es una sigla que significa Wide Area Information Systems. Es una herramienta que nos permite buscar información mas allá del título que esta posea. Esto significa que este sistema busca dentro del contenido del archivo mas que en su título.

A diferencia de sistemas como Gopher y Archie, que se plantearon a necesidades estudiantiles, Wais surge en un ambiente comercial con el objetivo de solucionar las necesidades de los ejecutivos de empresa. La idea fue de Brewster Kahle mientras trabajaba en Thinking Machines Corp., por el año 1991. Además formaron parte de este proyecto las empresas Apple Computers, KPGM y Dow Jones Co. Otra diferencia que podemos encontrar con Veronica y Archie es que Wais busca dentro de todo el texto mientras que los anteriores solo buscan cadenas de caracteres en los nombres de los directorio Gopher o Ftp respectivamente.

La característica principal de este sistema es que fue pensado para ser utilizado por personal no técnico por lo que debe ser fácil de usar y además de manejar diferentes tipos de información dispersas a través del mundo y administradas por distintos software.

Luego de que Wais fue creado, Kahle funda su propia empresa, Wais Inc. para enfocar sus esfuerzos en el mundo Wais.

La gama de información por la que se mueve es muy amplia y va desde documentos de física hasta software pasando por recetas de cocinas. La filosofía de Kahle es "Todo el mundo tiene algo que decir" . [Gil/94]

II. Arquitectura Wais

Se basa en el modelo cliente/servidor, donde los servidores son especializados en algún tema en particular. Aquí se ve la diferencia con el sistema Gopher en donde los servidores ofrecen una colección de información local y punteros a otros servidores. Existe también un servidor que contiene el directorio de servidores Wais disponibles, es la base de datos de bases de datos.

Los clientes Wais están capacitados para disparar consultas en paralelo a distintos servidores. Dada una consulta del usuario y las fuentes (servidores) donde este quiere buscar, el cliente genera en el formato preciso la consulta y espera que el servidor le envíe los documentos que podrían ser útiles para el usuario. El cliente toma estos documentos candidatos y los vuelve formatea para mostrárselo al usuario.

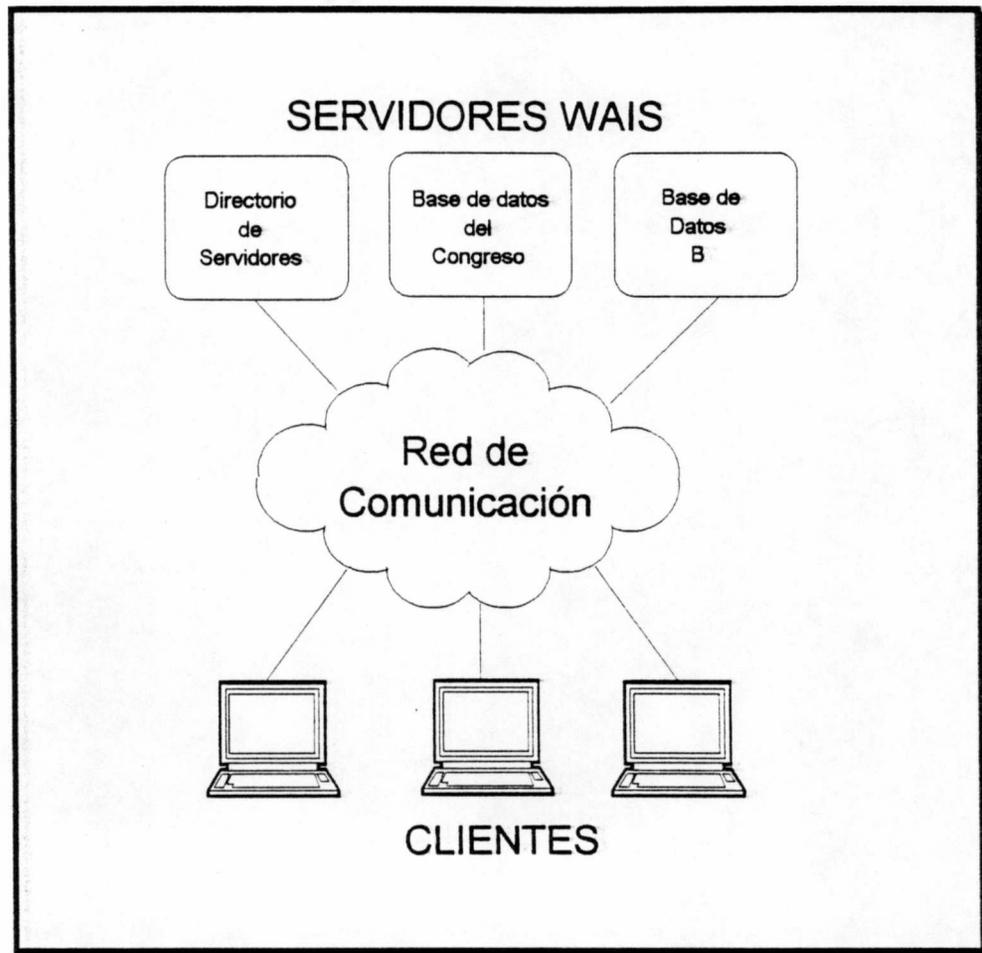


Fig A: Arquitectura Wais

La comunicación entre el cliente y el servidor se realiza mediante el protocolo Z39.50.

El Ansi Z39.50 es un estandar que define un protocolo dentro de la capa de aplicación del modelo de referencia OSI y se refiere particularmente a la recuperación de información almacenada en bases de datos. [Z39/91]

Este estandar describe el servicio de recuperación de información y especifica el protocolo de recuperación para la interconexión de sistemas abiertos. [Z39/91]

El servicio de recuperación de información especifica la actividad entre dos aplicaciones de diferentes computadoras: la aplicación iniciadora, origen y una aplicación que responde, destino. [Z39/91]

Este standar es usado particularmente en el área de las Ciencias de la Información y bibliotecas, y su objetivo es facilitar la interconexión abierta de usuarios de bases de datos con los proveedores de ellas. [Z39/91]

El protocolo de Wais esta basado en el Z39.50 y especifica los procedimientos y estructuras para el requerimiento de búsqueda (incluyendo la sintaxis), requerimiento para la transmisión de los registros de la base de datos localizados por una búsqueda, la respuesta al requerimiento, control de acceso y control de recursos. [DAV/90]

Las extensiones de Wais al estandar son primariamente soportar "relevance feedback". No se utiliza la facilidad para recuperación de texto propuesta por el Z39.50. Para recuperar un documento con el número xxx, la búsqueda se realiza con un especificación de consulta "Systems-Control-Number=xxx". Las extensiones de Wais también habilita al cliente a requerir una porción del texto del documento. Para ello se utilizan otros parámetros ("Chunk-Code"). La porción del documento que concuerda con los valores de estos parámetros será enviada. Supongamos los siguientes valores[DAV/90]:

"System-Control-Number=xxx AND Line>1000 And Line<=2000"

se enviaran las líneas 1001 hasta 2000 del documento xxx.

Este protocolo requiere que el sistema consultado envíe un identificador único (System-Control-Number) del documento en respuesta a una búsqueda. Estos identificadores son usados por el sistema que origina la búsqueda para especificar el documento cuando el usuario requiere verlo o para realizar una búsqueda con relevance feedback.

La recuperación de grandes documentos dependen de la habilidad de especificar un rango del texto a buscar. Este será especificado con una extensión llamada "Chunks". Existen tres tipos de chunks definidos actualmente: byte, línea y párrafo.

Las facilidades de la especificación Z39.50 que soporta este protocolo son dos:

- Inicialización: el cual permite al origen proponer valores para inicializar parámetros. Esta incluye un "Init APDU" y una "Init-response APDU"**
- Búsqueda: habilita al sistema origen a consultar una base de datos en el sistema destino y recibir información sobre los resultados de la consulta. Incluye "Search APDU" y un "Search-Response APDU". Esta se extendió para soportar un nuevo tipo de consulta: "relevance feedback".**

III. Como Trabaja

Veamos mas detenidamente el proceso entre el cliente y el servidor. Primero el usuario corre un cliente Wais que puede estar trabajando en un sistema operativo Unix, Dos, Macintosh u otro. Usando el software del cliente, el usuario arma una consulta especificando en que fuentes desea buscar y cuales son los términos de su consulta. Una vez que esta es creada, el cliente comienza una comunicación con los servidores requeridos y transmite el requerimiento.[Mor/94]

El servidor contesta la comunicación, analiza la consulta y busca en su base de datos. Este selecciona cuales de los documentos podrian ser útiles y le otorga un puntaje según el algoritmo que posea. Este puntaje va desde 1000 o sea la concordancia más alta descendiendo hasta la minima.[Mor/94]



BIBLIOTECA
FARMÁTICA
U.N.L.P.

Los documentos seleccionados son enviados al cliente quien los presenta al usuario en formato tipo periodico, dando un orden de importancia, presentandolos como libros en un estante o mostrandolo geograficamente las fuentes de documentos.[Kah/89]

Puede ser que los documentos enviados por los servidores no sean de utilidad para el usuario. Wais ofrece otra posibilidad, realizar una nueva búsqueda pero orientada a documentos que sean similares a los encontrados. Esta técnica es llamada "Relevance Feedback" y la veremos mas adelante.

IV. Como Acceder a Wais.

i) Wais via un software cliente:

Existen clientes para muchos de los sistemas operativos actuales, como por ejemplo para Unix tenemos swais (simple wais) y waissearch. para Microsoft Windows tenemos wwais, para Macintosh tenemos waisstation e Hyperwais. Tambien hay para Sun, VMS, etc.

ii) Wais via Gopher, WWW:[KOC/93]

Es facilmente accedido desde un numero de mediadores de información como por ejemplo Gopher, WWW.

iii) Wais via Telnet

El usuario puede realizar una conexión vía telnet a alguno de los servidores listados a continuación. Estas computadoras provee acceso a clientes wais y generalmente corren el swais. A continuación veamos una pequeña lista de sercidoes públicos:

<i>Máquina</i>	<i>Login</i>
<i>hub.nnsc.nsf.net</i>	<i>wais</i>
<i>kudzu.cnidr.org</i>	<i>wais</i>
<i>quake.think.com</i>	<i>wais</i>

sunsite.unc.edu	swais
-----------------	-------

V. Como realizar una búsqueda Wais.

Podemos dividir el proceso de búsqueda en dos partes:

1. Encontrar las fuentes apropiadas.

Wais necesita saber en donde buscar debido a que si no se acota el espacio de búsqueda, esta puede tomar un largo tiempo para que Wais encuentre lo pedido o hallar en un contexto equivocado lo solicitado por el usuario.

Lo primero que ve el usuario al ingresar al servicio Wais es un menú de las bases de datos del sistema, referido usualmente como fuentes (sources).[KOC/93]

Veamos como es una pantalla swais:

SWAIS		Source Selection	Sources:285
#	Server	Source	Cost
001:	[archie.au] aarnet-resource-guide	free
002:	[weeds.mgh.harvard.edu]	AAtDB	free
...			
098:	[quake.think.com] directory-of-servers	free
...			
Keywords:			
<space> select, w for keywords, arrows move, <return> searches			
q quits, or ?			

La columna con rotulada con "#" indica el numero de línea para la fuente; "server" es la dirección Internet del servidor donde reside la base de datos ; "source" es una breve descripción del contenido de la base de datos y "cost" es el costo para utilizarla.[KOC/93]

El usuario puede querer seleccionar directamente una fuente o buscar una apropiada a la consulta que desea realizar: Para seleccionar solo se debe mover el cursor hasta la opción deseada y presionar la barra de espacio o un "." Para buscar una fuente dentro del directorio de servidores, se debe seleccionar del menú de servidores la fuente "directory of servers" (en el ejemplo anterior se debe seleccionar la fuente número 098) y realizar un búsqueda.[KOC/93]

Muchas veces la parte mas difícil de usar el Wais es la selección de las fuentes a buscar. Podemos utilizar las siguientes ideas:[HAH/94]

i. Realizar una búsqueda previa en el directorio de servidores utilizando claves que describan la fuente. En esta búsqueda preliminar se debe ser mas general en la especificación de las claves que en la búsqueda de un ítem.

Veamos un ejemplo: Queremos encontrar un lugar que este asociado con cocos y abono, por lo tanto necesitamos algún tipo de referencia que concentre los hechos geográficos sobre el mundo. Una forma de buscar seria utilizar las claves "geografía" y "mundo".

ii. Hallar vía ftp anónimo un resumen de las fuentes de Wais. Este resumen contiene una breve descripción de todas las fuentes de Wais agrupada por categoría. Esta es la forma mas fácil y rápida ya que solo hay que buscar la categoría de la fuente y elegir la mas apropiada. Este archivo puede ser obtenido accediendo a la máquina quake.think.com al directorio /pub/directory-of-servers.

2. Buscar el ítem.

Luego de seleccionar las fuentes, el usuario debe moverse hasta la sección keywords e ingresar la o las claves.

Si la fuente que selecciono el usuario es ERIC:

110: [nic.sura.net] ERIC-archive free

111: [sun-wais.oit.unc.edu] eric-digest free

...

Keywords: reform

searching ERIC-archive.src....

El cliente swais establecerá una conexión al servidor wais localizado en *nic.sura.net* y pedirá a este que busque en la base de datos por cualquier documento que tenga la palabra "reform". Cuando la consulta es realizada el resultado se mostrará en pantalla de la siguiente forma:

SWAIS		Search Results		Items:
40				
#	Score	Source	Title	Lines
001:	[1000]	(ERIC-arch)	TITLE: Social Studies Curriculum reform	284
002:	[945]	(ERIC-arch)	TITLE: Físical Policy Issue ans School R	264
...
018	[222]	(ERIC-arch)	TITLE: School-to-Work Transition: ITs Ro...	251

<space> select, w for keywords, arrows move, s for sources, ? for help

Swais puede recuperar hasta un máximo de 40 documentos, otros clientes aceptan más o menos. [KOC/93].

Para poder ver un documento solo hace falta moverse con el cursor sobre el y presionar enter.[KOC/93]

Cuando el material es indexado, ciertas palabras pueden ser excluidas del proceso de indexación y de esta manera, no cumplir la función de claves. Por default, cualquier palabra que aparezca más de 20.000 veces es excluida. Por ejemplo, un archivo de descripción de software conteniendo 24.000 entradas, cada cual comenzando con la frase: "Nombre del software" podría excluir de su índice las palabras "software" y "Nombre". Similarmente, el material indexado a menudo excluyen las conjunciones ("but", "and", etc.), preposiciones ("in" "around", etc.), artículos ("a", "an", etc.) y un número de verbos comunes ("tell", "is", etc.). estas palabras que no son consideradas para el índice se llaman "stopwords".[KOC/93]

VI. Características sobre WAIS [KAH/89]

1. Como acceder a la información

Existen varias técnicas para acceder a la información. Una es buscar en una estructura arbolada. Quizas es la mas familiar para el usuario ya que aprende este concepto tempranamente al utilizar los directorios. Esta propuesta trabaja bien cuando el usuario conoce la organización de los datos. Además es computacionalmente eficiente siempre y cuando la cantidad de archivos almacenados no sea grande. Otra técnica es utilizar el concepto de hipertexto donde la información se conecta mediante lazos (ver WWW para mas detalle). Una técnica muy diferente es la llamada navegación por contenido. Este método permite viajar entre grandes colecciones de datos y buscar en ellas con solo la utilización de una consulta en lenguaje natural. La característica principal es que no es una expresión lógica compleja como las utilizadas en las bases de datos convencionales. Esta es la técnica utilizada por Wais.[Gil/94]

2. Relevance Feedback Relevance Feedback

Wais sufre de varios problemas en la búsqueda. Uno es que carece de sensibilidad conceptual , esto quiere decir que no hay forma de decirle a Wais que las palabras deben de ocurrir en un cierto orden o proeveerle alguna información sobre el contexto. Otro problema es que no hay forma de excluir partes de la fuente.

Para sobrellevar estos inconvenientes, Wais tiene una técnica llamada relevance feedback mediante la cual el usuario indica que quiere buscar otro documento parecido al que encontró en la consulta anterior. De esta manera, se puede ir orientando la búsqueda hacia documentos mas adecuados. En esta técnica, Wais toma de los documentos ofrecido como guías las palabras mas adecuadas para generar otra búsqueda en la misma o en otras fuentes.

3. Selección de los documentos mas adecuados. Selección de los documentos mas adecuados

Cada documento es clasificado según sea su concordancia con la consulta realizada por el usuario. Dependiendo del algoritmo de indexación del servidor, el puntaje dado al documento es basado en la posición del término en el documento, el número de veces que el término aparece en la base de datos y la longitud del documento.

En la versión original de Wais no se incluían operaciones lógicas por lo que se hacia complicado buscar cuando se necesitaban incluir en la consulta dos o mas términos. Por ejemplo, si queremos encontrar los documentos relacionados con redes de computadoras, podriamos generar la consulta "redes y computadoras"¹; el servidor buscaría las palabras "redes", "y" y "computadoras" y esto seguramente, nos daría documentos que no tendrian nada que ver con la consulta original debido a que podría encontrar documentos que tuvieran 1000 veces la palabra "redes" y ninguna la palabra "computadora" o 250 veces la palabra "redes" y 250 veces la palabra "computadora" y a pesar de ser un excelente candidato, Wais podría considerar mas relevante a la primera opción que a la segunda. .[Mor/94]

¹ En a realidad debemos realizar esta consulta en ingles pero por claridad preferi realizarla en español.

Posteriormente fue creado un algoritmo que permitía consultas utilizando expresiones lógicas. Este algoritmo se llamó IUBIO el cual contenía expresiones lógicas y luego aplico un esquema de pesos de la distribución original al conjunto resultado. .[Mor/94]

Mas tarde apareció otro algoritmo más inteligente que no solo incluyo búsqueda con caracteres lógicos sino que utilizó un thesaurus y un algoritmo de puntaje de documentos mas intenso. Despues de asignarle un peso a cada documento, se le otorga un puntaje de 1000 al mas relevante mientras que a los otros que tengan un peso mayor que cero se le asigna un puntaje proporcional.[Mor/94]

Cabe destacar que los servidores y clientes trabajan juntos para reportar la relevancia final de los elementos encontrados.

CAPITULO V

World Wide Web

World Wide Web o simplemente W³ o WWW, es un sistema que permite explorar, buscar y recuperar una gran gama de recursos de diferentes naturaleza como imagenes, texto, sonido y video dispersos en el Internet. Tiene la ventaja de no solo trabajar con información propia sino que también puede utilizar datos generados por otros sistemas como Gopher, Wais, Ftp entre otros.

Esta herramienta tiene dos grandes particularidades. La primera es que permite buscar sin restricciones dictadas por el programa, es decir el usuario decide cual es el próximo paso en su exploración. La búsqueda es no lineal esto quiere decir que nos podemos mover libremente saltando de un documento a otro y esto es gracias a una técnica llamada hipertexto y la cual veremos mas adelante. Para darnos una idea, los datos estan conectados como una telaraña dandonos una idea de que estos son continuos como si no existieran fronteras abstrayendonos de que tipo de recurso se trata. La otra particularidad, tal vez la que la hace mas revolucionaria, es que permite la combinación texto con sonido, gráficas y video mostrando una gran creatividad en el desarrollo de las pantallas.

La web fue originalmente desarrollada en Suiza, en el centro de investigaciones del CERN. El proyecto comienza por el año 1989 bajo la responsabilidad de Tim Berners-Lee[Gil/94]. La idea fue crear una forma para que los físicos del instituto pudieran compartir sus trabajos y usar la información de la comunidad. Luego esta se expandió dentro de Internet como un mecanismo general para acceder a información y servicios. [HAH/94]

I. Introducción

En el lenguaje de la web, un documento de hipertexto es todo archivo que contenga datos y posiblemente tenga punteros a otros documentos. El programa

que se utiliza para leer los datos o sea el cliente se llama explorador (ver la sección de arquitectura para mas detalle) y navegar significa seguir los punteros.[HAH/94]

No existe en teoría, ninguna razón, para restringir a datos de tipo texto. En el mundo de hipertexto usamos la palabra hipermedia para referirnos a documentos que puedan contener datos de diferentes clases (gráficos, sonido, animaciones).[HAH/94]

El objetivo de W^3 es combinar hipertexto y búsquedas de texto completo para que toda la información de Internet pueda ser entrelazada según las necesidades del usuario.

Las operaciones básicas para acceder a los datos son dos. Una es mediante la navegación que se realiza a través del seguimiento de los lazos. La otra es realizar una búsqueda para encontrar directamente los datos necesitados.

Hipertexto, ¿Qué es?

Es un texto el cual es accedido o almacenado en una estructura no lineal. Cada pieza del hipertexto (o nodo) es conectado a uno o mas piezas de hipertexto mediante punteros. Cuantos mas nodos y punteros son agregados a la estructura, los nodos y punteros forman una telaraña. [KOC/93]

La idea es muy simple y veamos para ello un ejemplo de la vida cotidiana, la estructura de un libro. Primeramente para encontrar algo dentro del libro es una buena idea mirar el índice y saltar a la hoja que este nos indique. También podriamos encontrar en una hoja una referencia al pié de página para agregar información como también aclaración de bibliografía. Estas referencias transforman al libro, que parece ser lineal, en una estructura compuesta por lazos a diferentes porciones de datos. Todos estos saltos a distintos lugares podrian verse como lazos de información. [Gil/94]

Podemos ver el caso contrario, los directorios. Estos están almacenados en una estructura jerárquica. Para obtener un cierto documento en un sistema de directorio se debe empezar por el raiz y recorrer a través de rutas de sub-directorio. Los sistemas jerárquicos como catálogos de librerías o directorio de

archivos son útiles para almacenar y recuperar información que es estática y/o fácil de clasificar. Pero estos sistemas son frágiles cuando tiene que tratar información que es dinámica, multidisciplinaria, difícil de clasificar o colmada de referencias a cosas sin relación aparente. [KOC/93]

URL: Universal Resource Locator.

El URL especifica donde la información referenciada esta localizada en el Internet. Es la forma como WWW denomina a los documentos y describe que hacer cuando un puntero es activado. De la perspectiva del usuario esto significa "cuando el puntero es activado, saltar al recurso especificado para más información". Para más información ver apéndice A.

II. El modelo de datos WWW[CAI/92]

El modelo de datos WWW usa los paradigmas de enlace de hipertexto y búsqueda por texto en una forma complementaria ya que ninguno puede reemplazar la funcionalidad del otro.

La vista del mundo W^3 es un conjunto de documentos referidos unos a otros mediante punteros. Por la semejanza con la construcción de una araña se le llamo Web (traducción telaraña). El usuario ve en la pantalla un documento con partes sensitivas representando los origenes de los punteros. [BER/92] El nivel superior de todos los sitios es llamado "Home del usuario" o "Home page". Cuando se invoca al cliente, este se conecta primero al "Home page" que esta especificado el cual puede ser local o remoto.

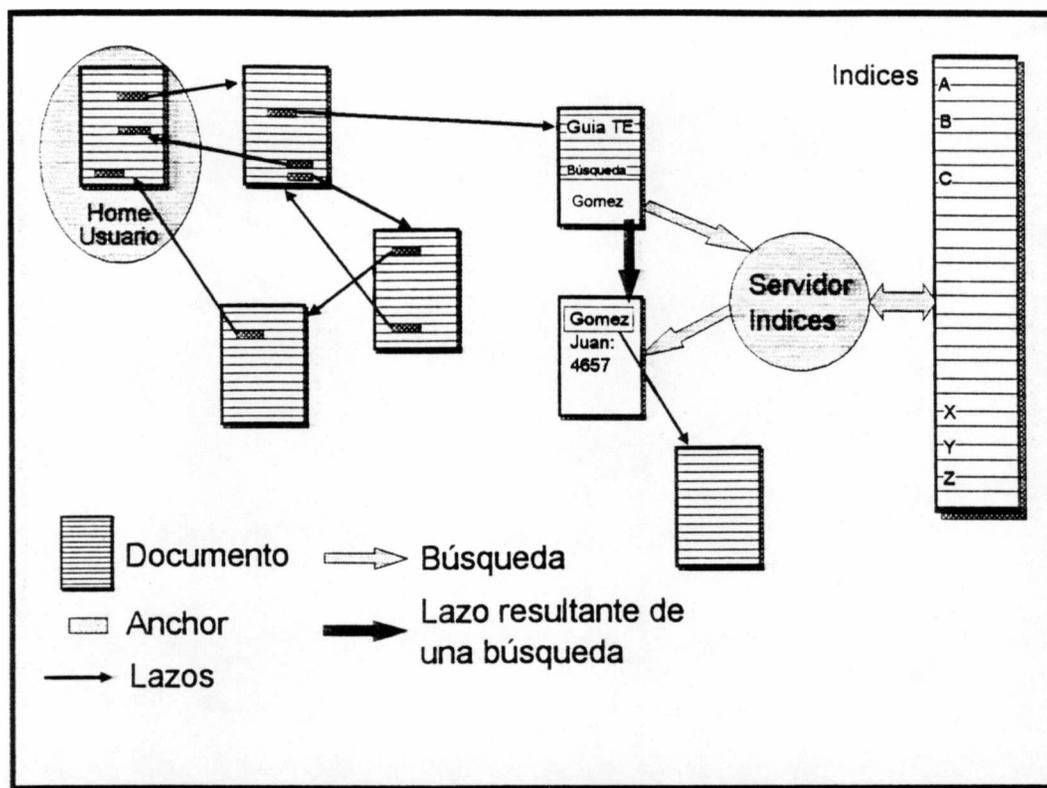


Fig A: Vista del mundo W^3 . [Dan]

El hipertexto por si solo no es práctico cuando tiene que manejar un gran conjunto de información contenidas en una base de datos. Agregar búsquedas al modelo de hipertexto dá a W^3 su poder. Los índices son documentos especiales que además de ser leídos pueden ser utilizados para hallar otros documentos. Para buscar en un índice, el lector da palabras claves (u otro criterio). El resultado de la búsqueda es otro documento (virtual) conteniendo punteros a los documentos encontrados. [BER/92]

Podemos ver varias características. La primera es que la información solo necesita ser representada una vez, ya que se hace una referencia y no una copia. La segunda característica es que los punteros permiten diagramar dinámicamente y sin restricciones la información, como lo hacemos naturalmente al pensar, resultando en diferentes estructuras de información en distintos estados de tiempo. Tercera característica, los índices son considerados documentos, de esta forma pueden ser encontrados por búsquedas y/o siguiendo los punteros. Un índice se representa como una sobretapa que describe los datos

indexados y las propiedades de la forma de búsqueda. Otra característica es que la web se extiende sin diferencias entre notas en una computadora personal a grandes bases de datos ubicadas en otros continentes. Por último, los documentos en la web no tienen por que existir, ellos pueden ser documentos virtuales generados por el servidor en respuesta a una consulta.

Un aspecto reconfortante y muy útil es que casi todos los sistemas de información existentes pueden ser representados en términos del modelo WWW. Un menú puede ser una página de hipertexto, con cada elementos enlazado a destinos distintos. La noción de nombres indexados en la red permite a un mecanismo de búsqueda dado y a bases de datos ser visibles con distintas direcciones, cada una representando opciones diferentes para el algoritmo de búsqueda.

III. Arquitectura WWW

El modelo utilizado es el cliente/servidor. En el mundo WWW el cliente es denominado explorador (browser) y es debido a la naturaleza de la web. Cada movimiento en la exploración requiere de una decisión de como será el próximo paso a seguir. Esto nos da una imagen de que la información es dinámica debido a que el usuario cambia virtualmente la organización de los datos en cada exploración. Otra de las particularidades que tienen estos clientes es que tratan con distintos protocolos, lo que lo hacen mas complejos que los vistos anteriormente. [Gil/94]

Existen diferentes tipos de exploradores y se diferencian en el conjunto de funciones que proveen al usuario. Por ejemplo, si estamos conectados directamente a Internet o utilizando una cuenta dial-up corriendo protocolos como SLIP o PPP¹, podemos utilizar exploradores que manejen sonido, gráficos, con complejos formatos de pantalla con una gran gama de imaginación para presentar lazos y combinaciones de los distintos recursos que dispone. Aquí comienza a jugar una nueva variable, la creatividad para presentar la información. Los requerimientos de hardware y software son mas complejos por lo que separamos a los exploradores como una manera de ingresar al mundo

¹ Ver glosario

WWW distinta de las vistas hasta ahora y las llamaremos interfaces gráficas de WWW. [Gil/94]

También podemos ingresar a WWW con un cliente que no necesite un equipamiento complejo pero lamentablemente no tendremos la potencia de las interfaces gráficas. En este caso deberemos de prescindir de sonido, gráficos y pantallas diseñadas complejamente, pero tendremos las funciones principales de WWW como explorar y buscar. Aquí encontramos dos tipos de exploradores los de caracteres y los de pantalla. Los de caracteres muestran los lazos mediante un número entre corchetes luego de la palabra que es origen de este lazo. Debajo de la pantalla nos dará un prompt para que escribamos la opción requerida ya sea una número o una selección de alguna función básica, como por ejemplo salir del programa, para indicar cual es nuestro próximo paso en la exploración. En los exploradores de pantalla podemos movernos libremente por la pantalla y los lazos están marcados con otro color. [Gil/94]

Los servidores aceptan la conexión, realiza lo pedido y le envía al cliente lo solicitado.

Protocolos Que Utiliza WWW [LEE/92]

El sistema WWW puede obtener información de muchas fuentes, usando protocolos existentes como por ejemplo, File transfer protocols - FTP, el protocolo de transferencia de archivos entre dos máquinas. El más utilizado es el ftp anónimo. WWW puede obtener información provistas por estos sitios. El FTP es definido en el RFC 959. Otro protocolo es el Network News Transfer Protocol - NNTP, permite el acceso a grupos y a artículos de news. Este protocolo permite un broadcast a una audiencia expandida. El Hypertext Transfer Protocol - HTTP, es un protocolo propio de WWW este para transferir información. Provee una función de búsqueda indexada. Usa el estilo normal internet sobre TCP/IP.

Veamos un ejemplo de como es el dialogo en el protocolo HTTP. La siguiente secuencia describe basicamente como el cliente requiere un documento de tipo hipertexto de un servidor HTTP dando una dirección de documento del tipo HTTP:

1. Conexión: El cliente realiza una conexión TCP/IP al host utilizando el nombre del dominio o el número IP con el número de port 80. El servidor acepta la conexión.²

2. Requerimiento: El cliente envía un requerimiento consistiendo de una línea de caracteres ASCII terminado por un <CR LF> . El requerimiento consiste de la palabra "GET", un espacio y la dirección del documento.

3. Respuestas: La respuesta a un simple "GET" es un mensaje en HTML. Este es una cadena de caracteres ASCII. Las líneas serán delimitadas con un retorno de carro opcional seguido por un caracter line feed. Las líneas pueden ser de cualquier longitud. Servidores bien configurados restringen las líneas a 80 caracteres incluyendo el par CR LF. El formato de los mensajes es HTML. Este permite que menús sean retornados como hipertexto.

El mensaje se termina con un cierre de conexión por el servidor. El cliente si esta bien configurado, leerá lo mas rápido posible. Este esperara a la acción del usuario (como por ejemplo la salida de la página) antes de terminar con la lectura del documento completo. El servidor puede imponer un tiempo máximo de espera, como por ejemplo 15 segundos.

Los errores son suministrados en texto que puede ser leído por el humano. No existe forma de distinguir una respuesta de error de una satisfactoria excepto en el contenido del texto.

4. Desconexión: La conexión TCP/IP es terminada cuando el servidor envió todo el documento. El cliente puede abortar la conexión cerrandola antes que termine la transferencia. El servidor, en este caso, no registrara ninguna condición de error. El servidor no necesita almacenar información sobre el requerimiento después de la desconexión.

²HTTP actualmente corre sobre TCP, pero podria correr sobre cualquier servicio orientado a conexión

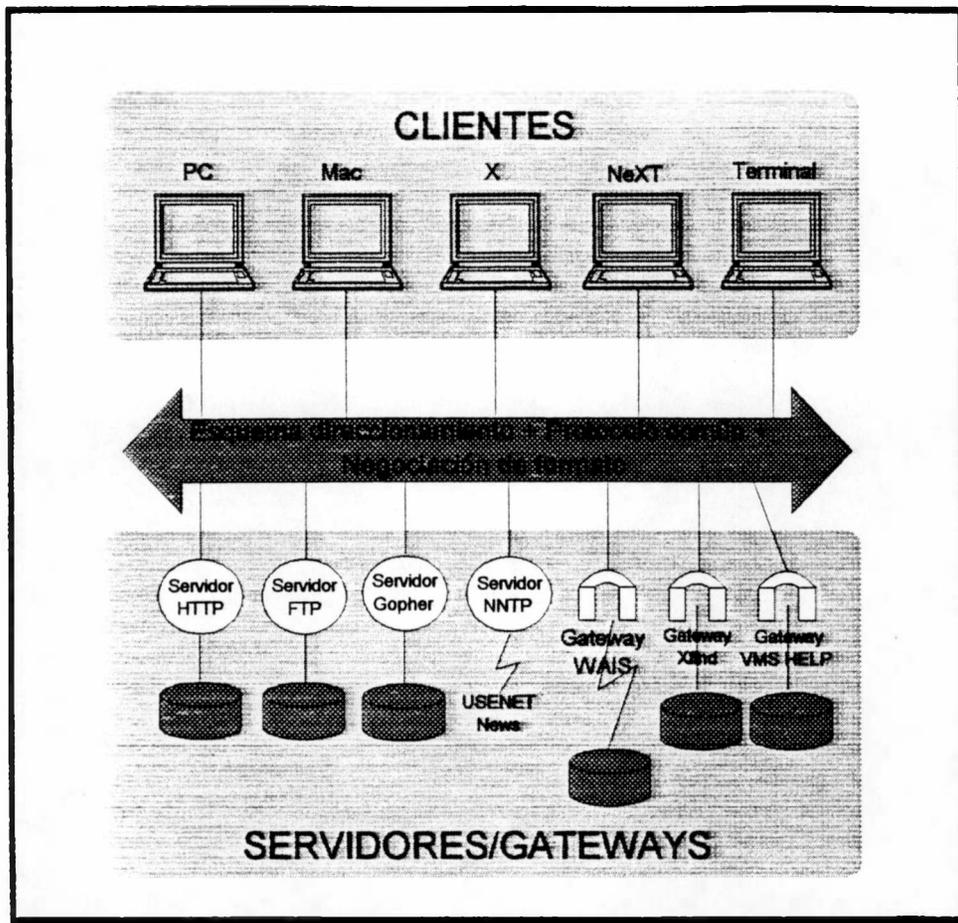


Fig C: Arquitectura WWW

IV. Como acceder a WWW.[Gil/94]

Tenemos 3 maneras de ingresar a WWW :

- **WWW vía software cliente:**

Cliente line browser es una versión de texto solamente para sistemas operativos con display solo caracter como Unix, DOS, o VM/CMS. En line browser se navega indicando a través de números de referencia, cual de los punteros se desea acceder.

No toma en cuenta que tipo de terminal se utiliza ya que todo lo que realiza es enviar una línea de texto despues de otra.[HAH/94]

Cliente screen mode browser : utiliza solo caracteres pero puede escribir en cualquier parte de la pantalla. es mas rapido y permite utilizar una interface un poco mas comp!aja. Necesita saber con que tipo de terminal esta trabajando el usuario.[HAH/94]

Para X-Windows, existen varios tipos de clientes. Se diferencian principalmente en el conjunto de herramientas utilizadas: como por ejemplo: el cliente ViolaWWW escrito y mantenido por Pei Wei de O'Reilly Associates. Tiene la facilidad de búsquedas, señalador e historia. El MidasWWW escrito por Tony Johnson de SLAC usando las herramientas Midas. El tkWWW escrito por Jospeh Wang del MIT. Esta basado en las herramientas "tk". El Erwise W³ fue escrito por estudiantes de la Universidad Técnica de Helsinki

- *WWW vía Telnet:[HAH/94]*

Si no se dispone de un cliente se puede utilizar un servicio público de WWW ejecutando un telnet a uno de los siguientes host:

<i>Máquina</i>	<i>Ubicación</i>
<i>info.cern.ch</i>	<i>Suiza</i>
<i>fserv.kfki.hu</i>	<i>Hungría</i>
<i>info.funet.fi</i>	<i>Finlandia</i>
<i>www.njit.edu</i>	<i>E.E.U.U.: New Jersey</i>
<i>ukanaix.cc.ukans.edu</i>	<i>E.E.U.U.: Kansas</i>
<i>fatty.law.cornell.edu</i>	<i>E.E.U.U.:New York</i>
<i>vms.huji.ac.il</i>	<i>Israel.</i>

Como login name debemos utilizar www. No todos estos host exigen el nombre de cuenta, algunos permiten el acceso automáticamente.

Los exploradores utilizados en host públicos de WWW son line browser o screen mode browser.

- **WWW vía una interface gráfica:**

El Mosaic fue escrito por Marc Andreessen del Centro Nacional de Aplicaciones para Supercomputadoras (NCSA) de la Universidad de Illinois, Urbana. Tiene una gran cantidad de ventajas. Utiliza diferentes colores para ayudar al usuario a mantener la pista de los punteros. Por ejemplo, Mosaic recuerda cualquier documento que el usuario haya leído. Cuando un enlace apunta a algún documento ya visto, Mosaic lo muestra con un color en particular. Esto evita perder tiempo sobre documentos ya leídos. Otra característica es que permite agregar notas personales dentro de cualquier documento. Estas aparecerán cada vez que el usuario lee el documento.[HAH/94]

V. Técnicas de Navegación y Herramientas[LEE/92]

La navegación es la esencia del hipertexto pero esta puede ser mejorada con un número de facilidades para hacerla mas eficiente y menos confusa.

1. Definir una estructura: A veces es bueno para el usuario tener la posibilidad de contar con una estructura de documentos construida, por el mismo autor del documento, con el objetivo de mejorar el entendimiento.

2. Mecanismo de historia: Este permite al usuario poder retroceder sobre sus pasos.

3. Índices: Un índice ayuda a nuevos usuarios a encontrar rápidamente en grandes bases de datos, nodos oscuros. Los esquemas con claves pueden ser incluidos en el tema general de índices. Este debe ser construido por el autor o automáticamente por el servidor, el explorador o un demonio. El índice esta compuesto de los títulos, una lista de claves, contenido del nodo o combinación de estos.

La operación especial que uno puede hacer con un sistema bueno de claves (que no puede hacer con un sistema normal de hipertexto) es hace una rápida búsqueda sobre claves múltiples. Esto debe proveer una extensión del esquema de navegación de hipertexto.

4. Menú de enlaces: La navegación regular puede hacerse con los puntos de anclaje resaltados o mediante un menú. Puede ser útil como una forma alternativa de navegación tener un menú de todos los enlaces de un nodo dado.

5. Una web de índices: En WWW, un índice es un documento como cualquier otro. Se puede construir un índice para cubrir un cierto dominio de información. Por ejemplo, el CERN posee un índice de los documentos que hay en el centro de computación. Hay un índice separado de la guía de teléfono. Los índices pueden ser construidos por el proveedor original de la información o por un tercero como un servicio agregado.

Los índices pueden apuntar a otros índices. Una búsqueda sobre un índice puede retornar como resultado otro índice.

VI. La vista del proveedor de información:[LEE/92]

Para los datos existentes, los exploradores WWW pueden acceder a través de protocolos FTP, NNTP o vía HTTP y un gateway. De esta forma la masa crítica de datos es rápidamente accedida.

Los autores pueden crear documentos simplemente creando archivos en texto plano, usando el lenguaje SGML o el editor de W³ y luego integrándolos a la web. Los punteros del hipertexto pueden hacerse a cualquier dato en cualquier servidor W³, ftp, Wais, Gopher o News ya que el cliente es el que tiene la habilidad de presentarle al usuario los datos como hipertexto. Corriendo los demonios de ftp o HTTP podemos poner los datos a disposición de la web.

En caso de sistemas de información existentes con gran cantidad de datos, uno consideraría la posibilidad de escribir un servidor para proveer una vista de hipertexto a los datos sin tocarlos o sin modificar los procedimientos utilizados para mantener la base de datos. Para ello se podría tomar como base un servidor ya hecho y modificarlo según las necesidades requeridas. Existen actualmente gateways para VMS/HELP, Hyper-G y Wais. (Ver Fig. B).

La ventaja principal de tener gateways es que se puede conservar los procedimientos para mantener y crear los datos.

En resumen:

- Si se tiene algunos archivos de texto plano: Se puede escribir rápidamente un archivo pequeño de hipertexto que apunte a ellos. Para hacerlos accesibles se puede utilizar demonios FTP anónimos o HTTP.

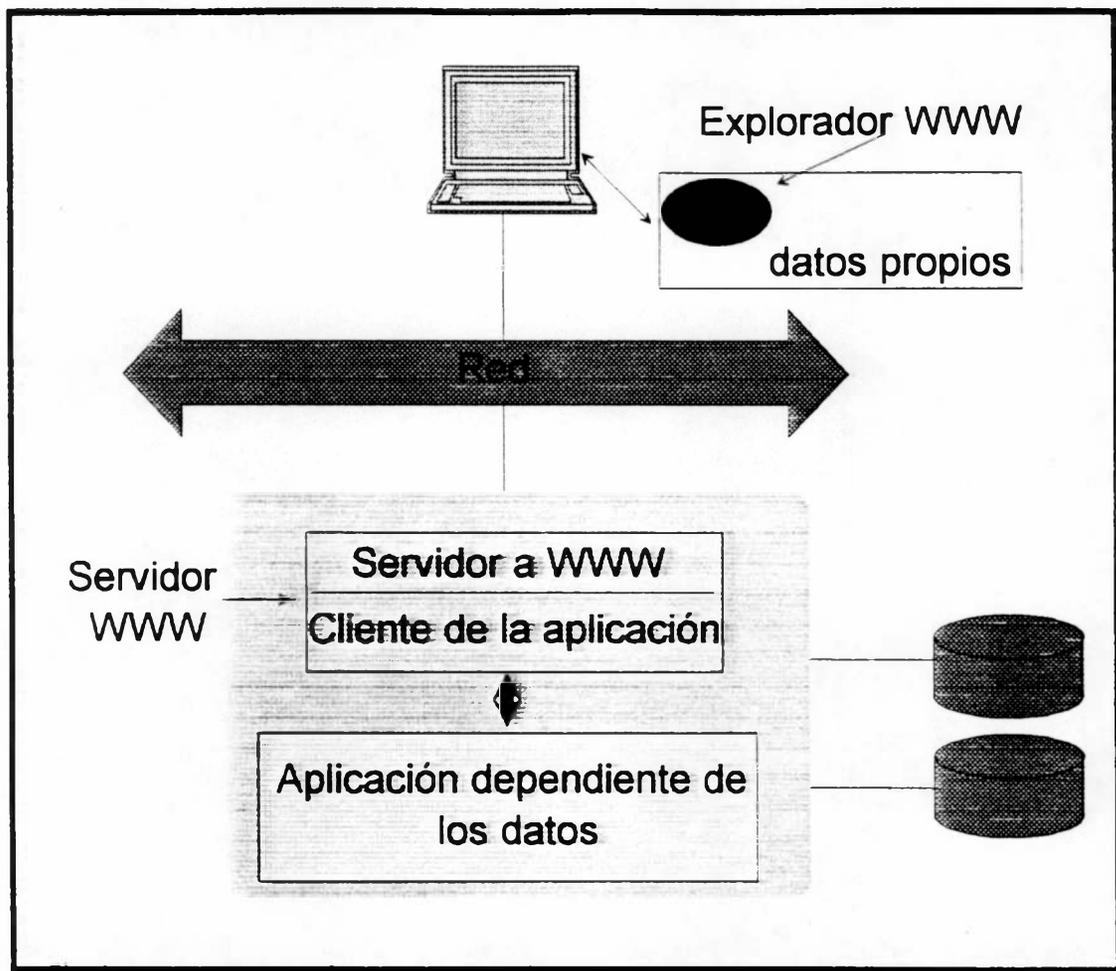


Fig B: Vista del proveedor de información.

- Si se tiene muchos archivos: En este caso para velocidad en el acceso, será mejor un demonio HTTP. El usuario puede escribir un árbol de hipertexto

en un enlazador de archivos HTML o puede generar una árbol automáticamente de la estructura arbolada de directorio. Si se quiere generar un índice de texto completo, se podría usar el software de dominio publico WAIS entonces los datos serán accesible (como texto plano, no como hipertexto) a través de un gateways WAIS.

• Si se tiene una base de datos existente: Un gateways W^3 puede correr sobre el sistema existente, poniendo la información disponible sin modificar los procedimientos que manejan las bases de datos. La forma en que trabajan estos gateways es basicamente la siguiente:

- Los menús son transformados en un conjunto de punteros**
- Las distintas opciones de búsqueda se transforman en documentos indices.**
- Los procedimientos usados para administrar y crear la información quedan inalterables.**

I.5 Servidores[LEE/92]:

Existen actualmente servidores W^3 para Unix, VMS, OS2, Windows NT¹ y deben ser configurados por los administradores de sistemas. Los servidores existentes incluyen aquellos para:

• Archivos: los servidores de archivos corren sobre Unix, VMS o VM para distribuir los archivos existentes a los exploradores de hipertexto. Los directorios del sistema de archivo son representados como listas de archivos de hipertexto de aquellos contenidos en los directorios. Los autores pueden ofrecer archivos de texto plano o vinculados mediante la técnica de hipertexto.

•VMS/Help: para información en formato VMS/Help, un servidor corre bajo VMS

¹ Datos obtenidos a diciembre de 1994.

- *Oracle: Un servidor Oracle genérico ha sido escrito por Arthue Secret para permitir el acceso a las bases de datos Oracle desde W³. Actualmente acepta sentencias SQL "select" para búsqueda y corre bajo Unix.*
- *GNU INFO: es un script perl que corre bajo Unix y provee el acceso a la base de datos Gnu Info de documentación online como hipertexto.*

VII. Relaciones entre WWW y los demás sistemas

Los servicios de Internet accesibles a través de W³ son: Wais, News, Gopher, Telnet, sitios ftp anónimos (Archie), X500, WHOIS, entre otros.[FAL/94]

El gateways WWW a archie y ftp:[FAL/94]

Se puede utilizar a WWW para realizar consultas archie y recobrar los archivos mediante ftp. WWW provee una elección de servidores hipertexto para archie. El resultado de una consulta es mostrado como un página de hipertexto que provee punteros a nombres de archivos que concuerdan con la clave, un enlace al directorio que contiene aquel archivo, un enlace al sitio que mantiene el sitio ftp en el cual el archivo fue encontrado. Seleccionando cualquiera de estos punteros se abrirá una sesión ftp al host seleccionado. Seleccionando el enlace al archivo se recuperara el archivo y se mostrara por pantalla. Seleccionando el enlace al directorio. se mostrara un menú con todos los archivos en ese directorio. Seleccionando el enlace al host. se abrirá una sesión ftp comenzando con el directorio ftp raíz de aquel host.

Wais y Web [CAI/93]:

Como vimos anteriormente, el protocolo Wais permite búsquedas basadas en el texto y recuperar luego de una búsqueda. Los índices en donde buscar son encontrados en un índice maestro. Se demostró que esta búsqueda en dos pasos es suficiente para cubrir el mundo de los datos en Wais. No existe, sin embargo, herramientas para navegar y mostrar al usuario los recursos disponible o guiarlo a través de los datos.

Sistemas de Menús y Web[CAI/92]

Los sistemas Gopher y Prospero usan el modelo directorio y archivo (o menú y documento) para implementar un sistema de información global. Este se puede traducir naturalmente a la web ya que cada directorio (o menú) esta representado por una lista de elementos enlazados a otros directorios o a archivos (documentos). Estos sistemas son muy confortables para lectores que están acostumbrados a sistemas de archivos jerárquicos, para quienes el concepto de directorio esta bien establecido.

Un ejemplo de transformación de un sistema de menú sobre la web es hecho por un cliente W³ que incorpora el protocolo Gopher y de alguna manera permite punteros a este sistema. Es verdad que un menú es necesariamente un medio mas restringido de comunicación que el hipertexto: una página de hipertexto puede acarrear mas información al usuario sobre las posibilidades a ser seguidas por utilizar un formato mas flexible.

Consideraciones Finales

Presente.

El propósito fundamental de este trabajo es dar a conocer algunas de las herramientas mas utilizadas hoy en día en la búsqueda y exploración en el Internet. Saber como encontrar la información es mucho mas importante que el conocimiento mismo.

Los cambios que estamos viviendo en este campo desde hace menos de 10 años son vertiginosos y revolucionarios. Cada día aparece algo nuevo, un software que tal vez, cuando el usuario termina de entenderlo pasa a ser algo obsoleto. Por lo que es imposible tener los datos actualizados al día .

Hace cuatro años atrás, comenzaron los esfuerzos para solucionar el problema de encontrar información en un medio donde los datos crecían exponencialmente, Internet y donde realmente era imposible hallar algo con una búsqueda manual. Varios grupos comenzaron a pensar como solucionar este problema y gracias a sus esfuerzos surgen herramientas con el objetivo principal, hacer viable la información. Cada una de las vistas en esta tesis tienen objetivos distintos y veamos en forma resumida cuales son:

<u>Clasificación</u>	<u>Herramienta</u>	<u>Objetivo</u>
<i>Básicas¹</i>	<i>Correo electrónico</i>	<i>Intercambio de mensajes entre usuarios</i>
	<i>Uso remoto (telnet)</i>	<i>Permite ingresar en una computadora remota.</i>
	<i>Transferencia de archivos (ftp)</i>	<i>Transfiere archivos entre dos computadoras</i>
<i>Exploración</i>	<i>Gopher</i>	<i>Servicio para la navegación de recursos de información</i>

¹ No desarrolladas en este trabajo. Ver glosario

		<i>almacenados en directorios en forma jerarquica.</i>
	<i>WWW</i>	<i>Servicio para la navegación de recursos en un ambiente hipermedia</i>
<i>Búsqueda de recursos</i>	<i>Archie</i>	<i>Servicio para búsqueda de archivos en sitios FTP anónimos.</i>
	<i>Veronica</i>	<i>Servicio que ayuda en la busqueda dentro de los titulos en los menus de Gopher.</i>
	<i>Judhead</i>	<i>Idem anterior pero se restringe a un sector de los menus de Gopher.</i>
	<i>Wais</i>	<i>Servicio que ayuda a buscar dentro de archivos.</i>
<i>Búsqueda de usuarios</i>	<i>Netfind</i>	<i>Servicios que ayudan a encontrar direcciones de usuarios</i>
	<i>Whois</i>	<i>Servicios que ayudan a encontrar direcciones de usuarios</i>
	<i>Knowbots</i>	<i>Servicios que ayudan a encontrar direcciones de usuarios</i>
	<i>CSO²</i>	<i>Servicio que permite buscar estudiantes o miembros de una casa de estudio.</i>
<i>Organización</i>	<i>Prospero</i>	<i>Sevicio que nos ayuda a organizar datos a traves de vistas virtuales.</i>

² No desarrollada. Ver Indice.



La pregunta natural que surge es ¿Cual es la mejor?. Cada una de ellas trabajan muy bien en el área para la cual ellas fueron diseñadas. Debido a la naturaleza del Internet, juntas tienen mayor poder.

Hoy Gopher, Veronica, Wais, Archie, WWW son herramientas poderosas que nos facilitan a encontrar datos dispersos alrededor del mundo. Estas herramientas están siendo mejoradas para lograr manejar los nuevos tipos de datos que están surgiendo y para satisfacer nuevas demandas de los usuarios. Todo parece que estas herramientas seguirán en el futuro pero no nos olvidemos lo cambiante de este campo el cual el día de mañana puede ofrecernos una nueva herramienta.

Futuro

Científicos e investigadores encuentran en el Internet como una fuente invaluable de información como así también una excelente vía para dirigir proyectos, intercambiar ideas e información. Internet es la base y el comienzo de una gran revolución que se nos avecina, la Autopista Informática (AI).

Todavía no existe una definición corta, consisa de que es la AI, tal vez porque es un concepto tan nuevo que todavía nadie sabe bien que es y a donde llegará. Pero si podemos decir cuales son sus componentes.

Las componentes de la AI son: un conjunto de redes públicas y privadas de alta velocidad, interactivas, conectadas a través de una tecnología satelital, terrestre con o sin cable que despacha información desde y hacia hogares, negocios e instituciones públicas y privadas. También la compone la información que se representa a través de bases de datos, cartas, documento, imágenes, sonido, filmaciones o software. La compone los dispositivos para acceder a esta información como televisores(interactivos), computadoras, radios, teléfonos y otros productos. Las compone también los usuarios que generan, proveen y administran esta información y también quienes la usan.

Gracias a todas estas componentes juntas, la AI nos da la sensación de una red sin fronteras, dinámica, invisible, sin límites, de mecanismos de transferencias, aplicación de la información contenido y seres humanos.

Según el sociólogo Alain Touraine, “las grandes industrias del porvenir son la educación, la información, la cultura, el conocimiento”. En un mundo tan cambiante, la educación es la preparación para estar aptos a los cambios que se vendrán. Muchas serán las ventajas que esta AI proporcionará a la enseñanza. La tecnología ayudará a profesores y alumnos a participar de clases que se proyectaran mas allá de las paredes del aula compartiendo experiencias e información con sus iguales y también con expertos alrededor del mundo sin moverse físicamente del salón de clases. El progreso no solo traerá “rosas”, incrementarán la droga y la delincuencia en las escuelas. La tecnología también podrá ayudar, permitirá monitorear las escuelas para prevenirlas de todo acto delictivo, (esto ya se está implementando en algunas escuelas en los EEUU). Ofrecerá a aquellos estudiantes que viven en zonas alejadas las mismas oportunidades que tienen los que viven en grandes ciudades. Cambiará la forma de educar ya que los libros pasarán a las computadoras y convertirá la educación en un entretenimiento. Aunque parezca increíble estas son las intenciones de Microsoft, el libro electrónico. Este no solo contará con el texto sino con sonido e imágenes y otras cosas más.

En el campo de la salud también habrá importantes cambios. La AI permitirá la utilización de la tecnología para compartir información, opiniones con especialistas de todo el mundo para el diagnóstico, tratamiento y seguimiento de pacientes. Juntar talentos y recursos sin importar en que parte del mundo estos se encuentran.

Abrirá las puertas para la creación de comunidades que surgirán sin el problema de fronteras ni distancias. Permitirá a aquellos que están lejos de su país poder compartir su cultura y sociedad.

También afectará nuestra vida cotidiana. Empresas como Microsoft y Oracle están desarrollando un nuevo sistema operativo para una nueva aplicación del sistema de cable llamado “video por demanda” en el cual el usuario podrá seleccionar directamente desde su televisor (televisión interactiva) que programa verá. La AI permitirá trabajar directamente desde el hogar realizando las mismas tareas sin movilizarse físicamente reduciendo costos y contaminación ambiental. Las compras se realizarán directamente desde los hogares. Un cambio total.

Mucho estan haciendo las grandes compañías internacionales de hardware, software y de servicios para la construcción de esta autopista. Según Bill Gates, dueño de la empresa Microsoft y un verdadero visionario, " La AI revolucionará al mundo aún mas que la invención del teléfono y cambiará para siempre la forma de la sociedad". Pero no solo Microsoft esta pensando en el futuro, empresas como Sony, IBM, Apple, Motorola, prestadoras de servicios telefónicos como AT&T y de cables , entre otras, estan enfocando sus esfuerzos en la participación de esta AI.

Incluso el gobierno de los EEUU esta poniendo atención a este nuevo movimiento, invertirá 2 billones de dolares en los próximos 5 años para el desarrollo de la AI. Pero que este gobierno este en esto quizas no sea novedad ya que por el año 1988, el actual Vicepresidente de los EEUU Al Gore, comenzó a hablar de la autopista y el gran potencial que esta tendrá en el siglo 21 ya que cambiará nuestras vidas. El tambien asegura que revitalizará la economía norteamericana luego de la gran caída que tuvo el mercado despues de la guerra de Vietnam. El Presidente de los EEUU no hubiera dado recursos si esta no tendrá grandes implicaciones en el futuro de Norteamérica.

Sin entrar en el mundo de la economía internacional ya que no es mi intención ni tampoco el tema de este trabajo, quiero notar algo que puede verse a simple vista y que puede ayudar al progreso de nuestro país. La Argentina se tendrá que poner a la par en tecnología de los países desarrollados para avanzar analizando mas de cerca que es lo que esta pasando en países desarrollados como EEUU en lo que se refiere a tecnologías e información. Surgirán nuevos sectores en la economía y con ellas nuevas oportunidades de trabajo. Si los expertos mas grandes estan viendo en las AI el futuro, nosotros no debemos resagarnos.

Hay muchas cosas mas que surgiran gracias a esto. Los especialistas predicen que los cambios antes del 2000 serán el uso de dispositivos de reconocimiento de voz que reemplazarán los teclados y paneles. Las computadoras tomarán mas características de la televisión a medida que la información multimedia aumente y sea mas sofisticadas y la televisión adquirirá

mas funciones de las computadoras al permitir seleccionar el contenido de la programación y comprar a traves de las pantalla.

Las predicciones a largo plazo son que toda la información pasará a ser digital. En algunos paises (como los EEUU) se crearán las comunidades virtuales, solo las personas se encontrarán cara a cara si es necesario. Los telefonos celulares continuaran pero no trabajando de la misma manera, se utilizará tecnologia satelital que permitira comunicarnos a traves del globo. La televisión interactiva tomará tiempo ya que es necesario que las compañías telefonicas y de cable hagan cambios de importancia. Y el desarrollo de la realidad virtual que hoy en día esta en sus comienzos.

Muchas son las razones para entrar en este movimiento del siglo 21. Nos dará mucha información, nos ampliará el conocimiento, pero esa nueva tecnologia nos exigirá un mayor grado de conocimiento para comprenderla y manejarla.

Pienso que la Argentina tendrá que estar preparada para este cambio y lo mas importante es que no debe quedar resagada a los adelantos tecnológicos. Lo principal es ingresar y ya dentro de este movimiento, podremos conocer inmediatamente cuales son lo avances que estan apareciendo en este campo.

APENDICE A

Como WWW denomina a los documentos

La forma en que WWW nombra a un documento, provee un método para que el cliente halle al servidor y este encuentre lo pedido. En el modelo WWW, un nombre también puede especificar una parte del documento a ser seleccionado por la aplicación. Es decir, concierne a la sintaxis que debe tener el nombre de un documento o parte de él (un punto de anclaje) para ser referenciado de cualquier parte del mundo.

Es probablemente uno de los aspectos mas cruciales de diseño y estandarización en un sistema de hipertexto.

Actualmente son varios los protocolos utilizados para recuperar la información por lo tanto la dirección debe ser capaz de aclarar muchos protocolos, métodos de accesos o esquemas.

Un enlace de hipertexto a un documento puede ser especificado usando un nombre lógico. Este es casi siempre, la única forma de evitar el problema de los documentos cuando son movidos físicamente.

Cualquier referencia permanente a un documento debe ser un nombre lógico mas que una dirección física. El cliente esta de alguna manera preparado para seguir varios pasos de la traducción antes de encontrar el documento final.

El esquema de WWW cumple con todos estos requerimientos pero es por otra parte abierto a agregar nuevos protocolos. Para este propósito un prefijo es usado para identificar el protocolo (y de alguna manera el esquema de nombramiento) para ser usado. Los clientes que no tiene tal protocolo en su repertorio deben acudir a un gateways para su traducción.

Esquemas de nombres de WWW

El formato de un nombre hipertexto consiste de:

- **el nombre de un subesquema de nombramiento a ser utilizado**
- **el nombre es un formato particular a ese subesquema.**
- **un identificador de desplazamiento particular dentro del documento (opcional)**

Por ejemplo, el formato para todo método de acceso basado en internet es:

subesquema://host.domain:port/ruta/ruta#desplazamiento

Un caracter "?" seguido de palabras separadas con signos "+" permite realizar una búsqueda en un índice

Ejemplos:

- **Nombre de archivo calificado, referenciando a un documento en el espacio de nombre de archivos de un nodo internet y un punto de anclaje imaginario 123 dentro de el**

file:/cernvax.cern.ch/usr/lib/WWW/default.html#123

- **El próximo ejemplo se refiere a un punto de anclaje "greg" en el mismo documento en el cual el nombre apareció anteriormente.**

#greg

1. Nombres de subesquemas de nombramiento:

Los diferentes esquemas usualmente utilizan diferentes protocolos en la red. El formato de la dirección después del nombre del esquema es una función del esquema en particular.

Subesquema	Descripción
<i>file</i>	<i>acceso a archivos, usando el explorador y/o gateways.</i>
<i>news</i>	<i>Acceso a artículos de news y a newsgroup, usando normalmente protocolo NNTP</i>
<i>http</i>	<i>acceso a cualquier información usando una búsqueda HTTP y protocolo de recuperación. El direccionamiento interno del sistema de información es mapeado a una ruta WWW.</i>
<i>telnet</i>	<i>acceso a una sesión telnet interactiva. esto es provisto solo como una interface a otro existente online que no puede ser mapeado sobre el espacio WWW.</i>
<i>gopher</i>	<i>el acceso se realiza utilizando protocolo gopher. Este protocolo es similar al HTTP pero usa un concepto de menús a archivo de texto en vez de hipertexto.</i>
<i>wais</i>	<i>El acceso es provisto usando la adaptación WAIS del protocolo Z39.50</i>
<i>X500</i>	<i>formato a ser definido.</i>

La dirección de un documento que podemos hallar en sistemas (tal como Wais) que no pueden ser actualmente accedidos directamente por WWW, pero sí a través de gateways, es codificada dentro de una de tipo HTTP en el gateways. Exploradores que no tienen la habilidad de usar ciertos protocolos pueden ser configurados automáticamente para usar gateways con ciertos esquemas de direccionamiento.

2. Direccionamiento para una búsqueda indexada.

Si un nodo de hipertexto es un índice o el servidor tiene uno asociado a él, la búsqueda puede ser hecha sobre el índice aclarando el nombre de este acompañado con una lista de claves, después del signo de pregunta:

<i>dirección-del-índice ? lista_de_claves</i>

donde:

la dirección del índice es una dirección de hipertexto normal.

En la lista_de_claves, claves distintas son separadas con el signo +

Normalmente, si la búsqueda fue exitosa el documento retornado contendrá puntos de anclaje a otros documentos que concuerdan con el criterio de selección.

El método de búsqueda, y las funciones lógicas y léxicas, pesos, etc. aplicadas a las claves dependerán de la dirección indexada. Un índice puede tener varias direcciones de hipertexto, las cuales se comportarán de diferente forma según sea la búsqueda. Por ejemplo: uno podría permitir una búsqueda sobre autores claves solamente, mientras que otra podría permitir una búsqueda de texto completo.

Ejemplo:

<http://cernvm/FIND/?sgml+cms>

3. Direcciones de archivos WWW

El formato de una referencia hipertexto de un archivo es una extensión del sistema de archivos Unix.

El formato es:

file://nodo/directorio/nombre

Los protocolos actuales usados por el cliente dependen de la implementación del explorador y el ambiente. Tipicamente, el explorador chequeará para ver si el nodo es local o uno en el cual los archivos están montados en alguna forma de sistema de archivo distribuido. Si el caso es ninguno de los anteriores entonces el explorador puede intentar con FTP anónimo u otros protocolos.

Veamos algunos ejemplos:

- **Un nombre de archivo totalmente calificado:**

**<A HREF="gopher://gopher.recyt.net/11/infogral/notidepo"... **

- **Una referencia relativa al mismo archivo:**

Referencias Historicas ...

** ..**

**Referencias Historicas .. **

¹ Ver el ejemplo completo en el apéndice C

En 1536, Don Pedro de Mendoza

• Una referencia relativa pero a otro archivo en el mismo directorio y en el mismo server:

**Barrios de San Telmo **

3.1 Accesos a directorios:

Si la dirección termina con un '/', el explorador recobrará el contenido del directorio especificado y generará una página virtual de hipertexto apuntando a su contenido. Además, este podría mostrar un archivo de información contenido en el directorio si alguno está presente.

4. Dirección de Hipertexto para News

Puede tomar cualquiera de los siguientes formas:

news:newsgroup	Lista de artículos disponible actualmente en el newsgroup dado. Este es una serie de caracteres alfanuméricos y puntos
news:*	Lista de newsgroup válido
news:id_mensaje	Artículo dado explícitamente. El id_mensaje es opcionalmente rodeado por corchetes angulares y debe contener un signo @.

Extensiones posibles de esto son los asteriscos para la lista de newsgroup. Este aumenta la carga con la lista completa.

No existe una forma de referenciar a artículos "no leídos". Mantener la pista de esto es el trabajo del explorador.

Ejemplos:

- **news:<1234678@cernvax.cern.ch>**
- **news:12345678@cernvax.cern.ch**

Ambas direcciones se refieren al mismo artículo

- **news:comp.sys.next.announce**

Este ejemplo se refiere a las lista de artículos en el newsgroup comp.sys.next.announce

5. Nombres relativos:

Las direcciones de un documento de hipertexto son normalmente dadas dentro del contexto de otro documento. Cuando la dirección de dos documentos son similares, el nombre relativo permite aclarar solo la diferencia entre ellos y de esta forma poder ahorrar espacio. Un ejemplo, es la dirección del destino de un enlace, que se especifica en forma relativa a la dirección del documento fuente.

Esto implica que ciertos caracteres, como por ejemplo "?", ".", tengan significados reservados para representar un espacio jerárquico y debe ser reconocido como tal, tanto en los clientes como en los servidores.

En el formato de direcciones WWW, las reglas para utilizar un nombre relativo son:

- **Si la parte de "esquema" es diferente, debe darse la dirección absoluta. De otra forma el esquema es omitido y:**

• Si la parte de "host" y/o "port" son diferentes, se debe dar el nombre del host y el resto de la dirección. El nombre del host puede darse usando la convención de Internet.

• Si la parte de acceso y host son las mismas, la ruta puede ser dada con la convención de Unix, incluyendo el uso de ".." para indicar el borrado de uno de los elementos de la ruta. Dentro de la ruta:

Si una barra está presente, la ruta es absoluta, si no

la última parte de la ruta de la dirección base (o sea el nombre de archivo del documento actual) se remueve y se añade a este lugar la dirección relativa.

En el resultado, toda ocurrencia "xxx/.." o"/." son removidas recursivamente, donde xxx es uno de los, elementos de la ruta (directorio).

El uso de barras "/" y doble puntos".." en este caso debe ser respetado por todos los servidores.

6. Direccionamiento HTTP:

La sintaxis de una dirección HTTP es:

<code>http://nombre_del_host[:port]/ruta[?palabras_para_busqueda]</code>
--

Ejemplo:

- `http://info.cern.ch/hipertext/WWW/Theproject.html`
- `http://crnvmc.cern.ch/FIND?sgml+examples`

²los [] significa que los parámetros dentro de ellos son opcionales.

Las direcciones HTTP encuadran con las convenciones WWW, incluyendo la posibilidad de usar el formato de búsqueda.

nombre del host	este es el nombre del servidor en la forma internet. Es preferible una forma numérica (como por ejemplo 128.141.201.74) en vez del nombre del domain name (o sea info.cern.ch). El nombre del host es mandatorio.
port	número del port. Puede usarse el nombre del servicio. No existe un lugar centralizado para nombres de servicios, ellos son definidos localmente por cada host. El número default es 80.

7. Direccionamiento Telnet

Es un caso especial de dirección WWW. Cuando se usa un direccionamiento telnet, la información solo puede ser recobrada usando una sesión interactiva. Esta tiene la desventaja que la información no puede ser automáticamente indexada o buscada. Este es usado para permitir un puntero a los sistemas de información que todavía no pueden ser introducido correctamente en la web.

La sintaxis es :

```
telnet: // [usr@] host [ : port ]
```

donde:

usr	es el nombre opcional del usuario utilizado para el logoneo. Si el usuario es omitido, se debe sacar el signo "@"
host	Este es el nombre del servidor en la forma internet. Puede usarse una forma numérica en vez de la forma domain name. El host es mandatorio.
port	Es el número del port. Si se omite el número del port debe ser omitido también el signo :. En este caso se utilizara el default 23.

Ejemplos:

telnet://www@info.cern.ch:23

telnet://www@info.cern.ch

telnet://info@cern.ch

8. Direccionamiento Gopher

Direcciones gopher indican que debe utilizarse el protocolo gopher para acceder a la información. El protocolo Gopher es un simple protocolo de Internet similar al HTTP. Este permite transferencia de menús o archivos planos³.

La sintaxis es, utilizando los [] como opcionales, :

gopher://nombre_del_host[:port][/tipo_g/[selector]][?búsqueda]

donde:

³HTTP expresa a los menús y archivos de texto como casos especiales de archivos de hipertextos.

<i>nombre_del_host</i>	<i>Nombre del host en forma de domain name o número internet. Es obligatorio.</i>
<i>port</i>	<i>Port numérico. Si se omite se asume 70 y debe omitirse también el ":".</i>
<i>tipo_g</i>	<i>Es un ítem gopher de tipo numérico del tipo ASCII permitido. Caracter "0" indica archivo de texto; "1" indica menú; "7" indica índice; "8" podría no usarse en WWW, usado para telnet, etc. El resto de la ruta es un selector gopher de tipo string.</i>
<i>selector</i>	<i>Este es el string enviado al servidor gopher para identificar la información requerida</i>

9. Direccionamiento WWW para servidores WAIS

Los servidores que utilizan protocolo WAIS pueden ser accedidos, como parte de la web, usando direcciones de la forma:

wais: // hostport / base_de_datos...

El acceso se realiza a través de un gateways que almacena los "archivos fuentes" conteniendo la descripción de los servidores WAIS. Esta dirección corresponde a la de un índice.

Actualmente las direcciones a documentos que quieren recobrase por un cliente requiere de la siguiente información, que es provista por el servidor en una lista.

<i>formato de documento</i>	<i>Este es normalmente "TEXT" pero existen otros formatos como por ejemplo PS, GIF</i>
<i>largo del documento</i>	<i>Es necesario para que el cliente pueda recuperar el documento en pedazos..</i>
<i>identificador del documento</i>	<i>Es una entidad consistiendo de campos numéricos. La representación binaria utilizada por WAIS es transformada para poder leerse en secuencias de campos, cada uno conteniendo un valor decimal, un signo de igual "=", un campo valor y un ";". Dentro del campo valor, hay que escapar si existen caracteres ilegales.</i>

10. Sintaxis de direccionamiento WWW: BNF

Veamos la sintaxis de las direcciones WWW en BNF (Backus Naus Form).

Se utilizan el signo "-" para indicar alternativas, y [] para indicar partes opcionales. Los espacios son solo para lograr una mejor representación en este documento, no se permiten espacios en una dirección W³.

Una dirección absoluta especificada en un enlace es un dirección_pto_ancla. La dirección pasada a un servidor es una dirección_doc.

<i>dirección_pto_ancla</i>	<i>dirección_doc[#punto_de_anclaje]</i>
<i>dirección_doc</i>	<i>dirección_http - dirección_archivo - dirección_news - dirección_telnet - dirección_gopher - dirección_wais</i>

<i>dirección_http</i>	<i>http://hostport/[ruta][?busqueda]</i>
<i>dirección_archivo</i>	<i>file://host/ruta</i>
<i>dirección_news</i>	<i>news:parte_grupo</i>
<i>parte_grupo</i>	<i>* - grupo - artículo</i>
<i>grupo</i>	<i>ialfa[.grupo]</i>
<i>artículo</i>	<i>xalfas@host</i>
<i>dirección_telnet</i>	<i>telnet://[usr@]hostport</i>
<i>dirección_wais</i>	<i>indice_wais - doc_wais</i>
<i>indice_wais</i>	<i>wais://hostport/base_de_datos[?busqueda]</i>
<i>doc_wais</i>	<i>wais://hostport/base_de_datos/tipow/digitos /ruta</i>
<i>base_de_datos</i>	<i>xalfas</i>
<i>tipow</i>	<i>xalfas</i>
<i>dirección_gopher</i>	<i>gopher://hostport/[tipo_g/[selector]][?busqueda]</i>
<i>hostport</i>	<i>host[:port]</i>
<i>host</i>	<i>nombre_del_host - número_del_host</i>
<i>nombre_del_host</i>	<i>ialfa.nombre_del_host</i>
<i>número_del_host</i>	<i>digitos.digitos.digitos.digitos</i>
<i>port</i>	<i>digitos</i>
<i>selector</i>	<i>ruta</i>
<i>ruta</i>	<i>void - xalfas[/ruta]</i>

<i>búsqueda</i>	<i>xalfas - [+búsqueda]</i>
<i>usr</i>	<i>xalfas</i>
<i>punto_de_anclaje</i>	<i>xalfas</i>
<i>tipo_g</i>	<i>xalfa</i>
<i>xalfas</i>	<i>xalfa[xalfas]</i>
<i>xalfa</i>	<i>alfa - \$ - _ - @ - ! - % - ^ - & - * - (-) - . - digito</i>
<i>ialfa</i>	<i>alfa[xalfas]</i>
<i>alfa</i>	<i>a - b - ... - z - A - B - ... - Z</i>
<i>digitos</i>	<i>digito[digitos]</i>
<i>digito</i>	<i>0 - 1 - ... - 9</i>
<i>alfanum</i>	<i>alfa - digito</i>
<i>alfanums</i>	<i>alfanum[alfanums]</i>

11. Evitando caracteres ilegales

La sintaxis de W^3 permite que una ruta contenga la mayoría de los caracteres ASCII, pero algunos de los caracteres utilizados para puntuación están excluidos. Las direcciones W^3 , a veces, son usadas para representar direcciones en otro espacio. Por ejemplo, cuando un servidor HTTP, usa nombres de archivos como nombres de documentos, o cuando las direcciones de algún otro protocolo (Gopher, Wais, etc.) son traducidas al formato utilizado en WWW.

En estos casos se usa una convención para poder traducir caracteres ilegales de estos nombres "foráneos" sobre el conjunto de caracteres legales.

La convención es que un signo de porcentaje (%) seguido de dos dígitos hexadecimales (0... 9 - a... f) se colocan en lugar del carácter utilizando su código ASCII hexadecimal.

Un ejemplo, si es necesario incluir un espacio un espacio, se reemplaza el mismo por %20 ó si nos encontramos con un signo de porcentaje este se traduce de la siguiente manera: el código ASCII hexadecimal del "%" es el 25 entonces se pasa a %25.

APENDICE B

Lenguaje HTML[LEE/93]

El sistema W³ define una sintaxis concreta en SGML para hipertexto básico usado para menús, resultados de búsquedas y documentación de hipertexto en línea. [CAI/92]

El lenguaje HyperText Markup Language (HTML) está definido en términos del Standard Generalized Markup Language (SGML) de la ISO. El SGML es un sistema para definir documentos estructurados y un lenguaje de señalización para representar instancias de aquellos documentos.

Cualquier documento SGML tiene 3 partes:

- 1. Una declaración SGML: fija las cantidades a procesar y define la sintaxis de los nombres de los tokens para especificar los valores. Por ejemplo, la declaración SGML en el DTD (Document Type Definition) HTML especifica que la cadena de caracteres que abre un rótulo es </ y el máximo de caracteres para un nombre es 40.**
- 2. Un prólogo incluyendo una o más declaraciones del tipo de documento que especifica los tipos de elementos, relaciones entre elementos y atributos, y referencias que pueden ser representadas por la señalización. El HTML DTD especifica, por ejemplo, que el elemento HEAD contiene a lo sumo un elemento TITLE.**
- 3. Una instancia que contiene el dato y la marca del documento.**

Usamos el termino **HTML** para referirnos al tipo de documento y al lenguaje de señalización para representar instancias de aquel tipo de documento.

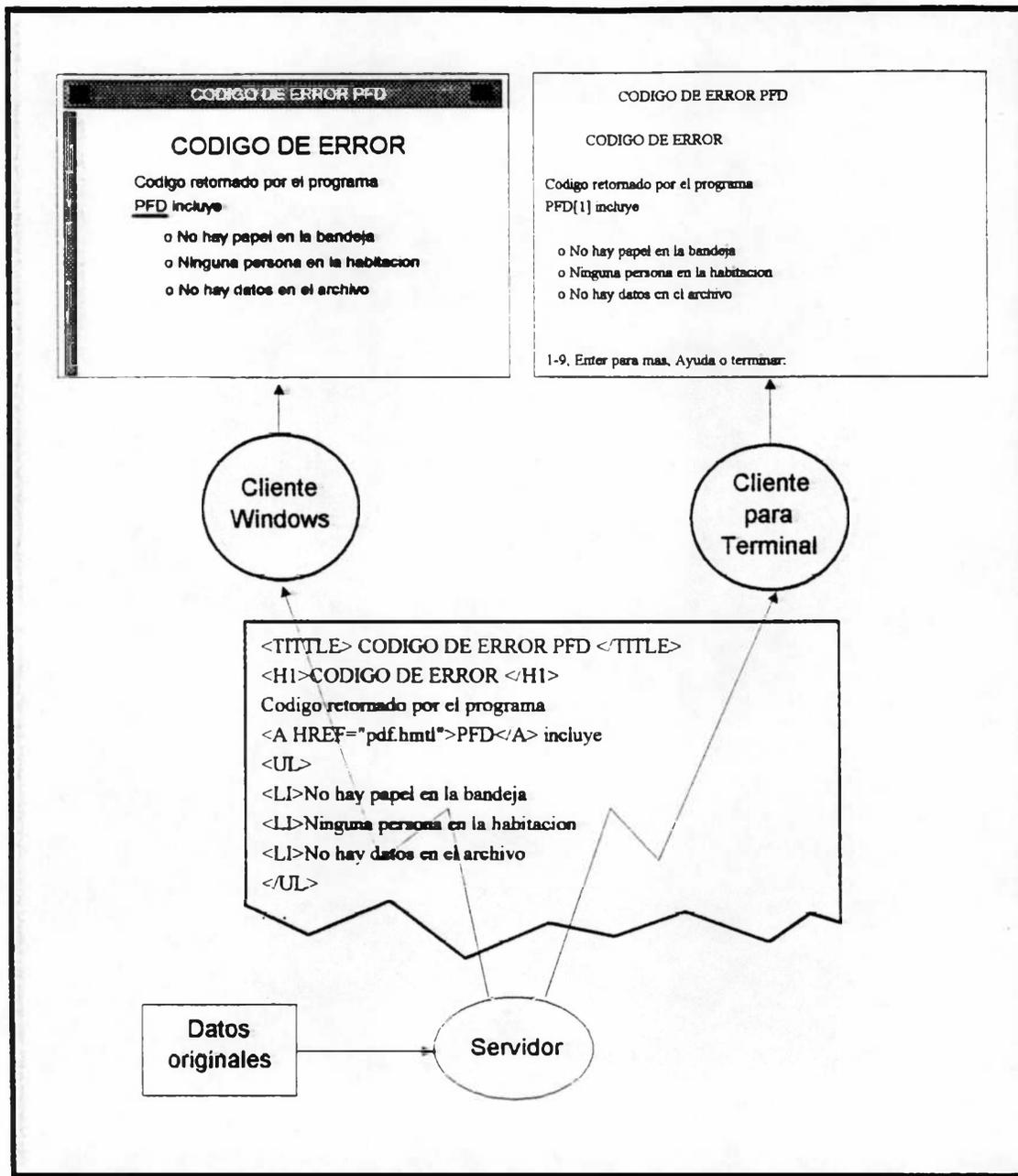


Fig A: Ilustración de los datos de hipertexto codificado.

1. Introducción

1.1 Texto estructurado

Una instancia HTML es como un archivo de texto, excepto por algunos caracteres que son interpretados como marcas. Estas marcas dan la estructura al documento.

La instancia representa una jerarquía de elementos. Cada elemento tiene un nombre, algunos atributos y algún contenido. Los elementos se representan con un rótulo de comienzo, que le da el nombre y atributos, seguido del contenido y terminando con un rótulo de finalización

Ejemplo

```
<HTML>  
  <TITLE>  
  Un ejemplo de una instancia HTML  
  </TITLE>  
  <H1>  
  Un ejemplo de estructura  
  </H1>  
  Aquí hay un párrafo típico  
  <P>  
  <UL>  
    <LI>  
    El Item uno tiene un  
    < A NOMBRE ="punto de anclaje">  
    punto de anclaje  
    <A>  
    <LI>  
    Aquí esta el item 2  
  </UL>  
</HTML>
```

Alguno elementos son vacíos como por ejemplo P y LI. Para el resto de los elementos, el contenido es una secuencia de datos caracteres y elementos anidados.

1.1.1 Rótulos:

Todo elemento (ver mas adelante) comienza con un rótulo y todo elemento no vacío finaliza con uno.

Cada rótulo de comienzo esta encerrado entre < y >. Un rótulo de finalización esta encerrado entre </ y >

Algunos rótulos tienen parámetros, llamados atributos. Ciertos atributos toman un efecto simplemente por su presencia otros son seguidos por un signo = y un valor. No importan si están en mayúscula, minúscula o mezclados. Por convención se utiliza las mayúsculas.

1.1.2 Tipo de elemento:

El nombre en un rótulo esta referido a un tipo de elemento declarado en el HTML DTD. Una declaración de tipo de elemento asocia a un nombre de elemento con :

- **Una lista de atributos con sus tipos y estados**
- **Un tipo de contenido que determina la sintaxis del contenido del elemento (uno de EMPTY, CDATA, RCDATA, MIXED o ELEMENT)**
- **Un modelo del contenido que especifica la forma en el cual los elementos y datos están anidados.**

Veamos los tipos de contenido:

Vacío: elementos vacío tiene la palabra clave EMPTY en su declaración.

Ejemplo:

<!ELEMENT NEXTID - O EMPTY>
<!ATTLIST NEXTID N NUMBER #REQUIRED>

Esto significa que:

<nextid n="27"> es legal

<nextid> es ilegal

<nextid n="abc"> es ilegal

Caracter: indicado mediante la palabra clave CDATA, indica que el elemento es dato de tipo caracter. El dato caracter es todo el texto hasta el próximo rótulo de finalización-en-contexto que es seguido de una letra.

Ejemplo:

<!ELEMENT XMP - - CDATA>

Especifica que el siguiente texto es legal:

```
<xmp>  
Este es un ejemplo. Parece  
que tiene <rótulos> y <!--comentarios-->  
en el, pero no es así. Aun esto  
</ es dato.  
</xmp>
```

Veamos un texto ilegal:

```
<xmp>  
No existe forma de representar  
un rótulo de </fin> en CDATA  
</xmp>
```

Este es debido a que una vez que se reconoce al rotulo completo (</fin>). La diferencia con el anterior es que este tenia solo el comiemo del rótulo.

Dato Caracter reemplazable:

Los elementos con un contenido RCDATA se comporta igual que el caso anterior excepto para referencias de tipo caracter y para referencias de tipo entidad.

Los elementos declarados como <|ELEMENT TITLE - - RCDATA> pueden tener cualquier secuencia de caracteres en su contenido.

° **Referencia caracter:** Para representar un caracter que podría ser reconocido como uno especial, como por ejemplo el <, usar una referencia del caracter. La cadena de caracteres &# señala una referencia caracter cuando es seguida de una letra o dígito. El delimitador es seguido del número decimal del caracter y un ;.

Ejemplo:

El número decimal del < es 60 =>

<title>

**Se puede representar en un RCDATA
un rótulo <fin>**

</title>

° **Referencia entidad:** El HTML DTD declara entidades para el <, >, & y cada uno de los caracteres ISO Latin 1 entonces nos podemos referir a ellos mediante nombres en vez de números.

Ejemplo:

Dentro de la lista de entidades la ö se representa ö

Kurt Gödel fue un famoso matemático.



Contenido elemento

Algunos elementos tienen, en vez de claves que dicen de que tipo es, un modelo del contenido que nos dice que modelo de datos y elementos anidados son permitidos. Si el modelo del contenido no incluye el símbolo #PCDATA, el contenido es de tipo elemento.

Los espacios en blanco son considerados señalización e ignorados. Cualquier caracter que no sea señalización, o sea caracteres de datos, son ilegales.

Por ejemplo:

<|ELEMENT HEAD - - (TITLE? & ISINDEX? & NEXTID? & LINK*)>

Declara que un elemento puede se usado como:

```
<head>  
  <isindex>  
  <title>  
    Ejemplo HEAD  
  </title>  
</head>
```

Pero lo siguiente es ilegal:

```
<head>  
  No esta permitido  
  datos aqui  
</head>
```

```
<head>  
  <isindex>  
  <title>  
    dos rótulos de isindex  
  </title>
```

```
<isindex>
</head>
```

Contenido mixto

Si el modelo de contenido incluye el símbolo #PCDATA, el contenido del elemento es parseado como un contenido mixto.

Por ejemplo:

```
<ELEMENT PRE - - (#PCDATA | A | B | I | U | P) +>

<!ATTLIST PRE
  WIDTH NUMBER #IMPLIED>
```

Esto dice que el elemento PRE contiene uno o mas elementos A,B,I,U o P o datos caracter.

Un ejemplo del elemento PRE:

```
<pre>
  <b> NAME </b>
    cat -- concatenado <a href="terms.html#file">archivos</a>
  <b> ejemplos </b>
    cat <xyz
</pre>
```

El contenido del elemento pre es:

- ° elementos A B
- ° la cadena de caracteres " cat - - concatenado"
- ° un elemento A
- ° la cadena de caracteres "\n"
- ° otro elemento B

- ° una cadena de caracteres "\ncat<xyz"

1.1.3 Comentarios y otras señalizaciones

Para incluir comentarios en un documento HTML, debe estar rodeado con <!-- y -->. Todo lo que este dentro de estos delimitadores será ignorado.

<? Instrucción de procesamiento

<! Declaración de señal. se termina con >

1.1.4 Líneas separadas

El caracter de separación se ignora si la primera pieza o la ultima pieza del contenido en un elemento

Ejemplo:

<pre> algo de texto pata ejemplo </pre>

es lo mismo que:

<pre>

algo de texto para ejemplo

</pre>

también se ignoran las líneas en blanco.

2. Elementos HTML

Se presentará una lista de los elementos mas usados en el lenguaje HTML. Los documentos debería contener un elemento inicial HEAD seguido por un elemento BODY.

2.1 Propiedades del documento entero:

Las propiedades del documento son definidos por los siguientes elementos. Ellos deberían aparecer dentro del elemento HEAD.

Elemento	Descripción
TITLE	El título del documento
ISINDEX	Este elemento informa al lector que el documento es un documento índice.
NEXTIND	Un parámetro enviado por editores para generar identificadores únicos
LINK	Relación entre un documento y otro. Un documento puede tener varios elementos LINK.

2.2 Formateo de Texto

Estos elementos los podemos encontrar en el elemento BODY de un documento. El orden en que ellos están descriptos es el lógico en el cual ellos deberían ser enviados al dispositivo de salida.

Hi	<p><i>Diferentes niveles de encabezamiento son permitidos. Uso típico de esto es, H1: encabezamiento con tipo de letra grande centrada, Una o dos líneas de espacio entre esto y lo siguiente.</i></p> <p><i>H2: Letra grande. Indentada a margen izquierda, etc.</i></p>
A	<p><i>Sección de texto que conforma el comienzo y/o final de un enlace de hipertexto. Son llamados "punto de anclaje". Los atributos de un rótulo de tipo punto de anclaje puede ser: HREF para marcar que es el comienzo de un enlace y por lo tanto es un texto sensitivo; NAME permite al punto de anclaje ser el final de un enlace; etc.</i></p> <p><i>Ejemplo:</i></p> <p><i>Un crimen serio es el que esta asociado con encarcelamiento</i></p> <p><i>La organización puede despedir a un empleado si este es convicto de un crimen serio</i></p>
P	<p><i>El elemento P marca el corte entre dos párrafos.</i></p>
ADDRESS	<p><i>Utilizado para información de dirección, autor, etc.</i></p>
BLOCKQUOTE	<p><i>Permite encerrar entre comillas texto de otra fuente para ser presentado especialmente.</i></p>

3. Ejemplo

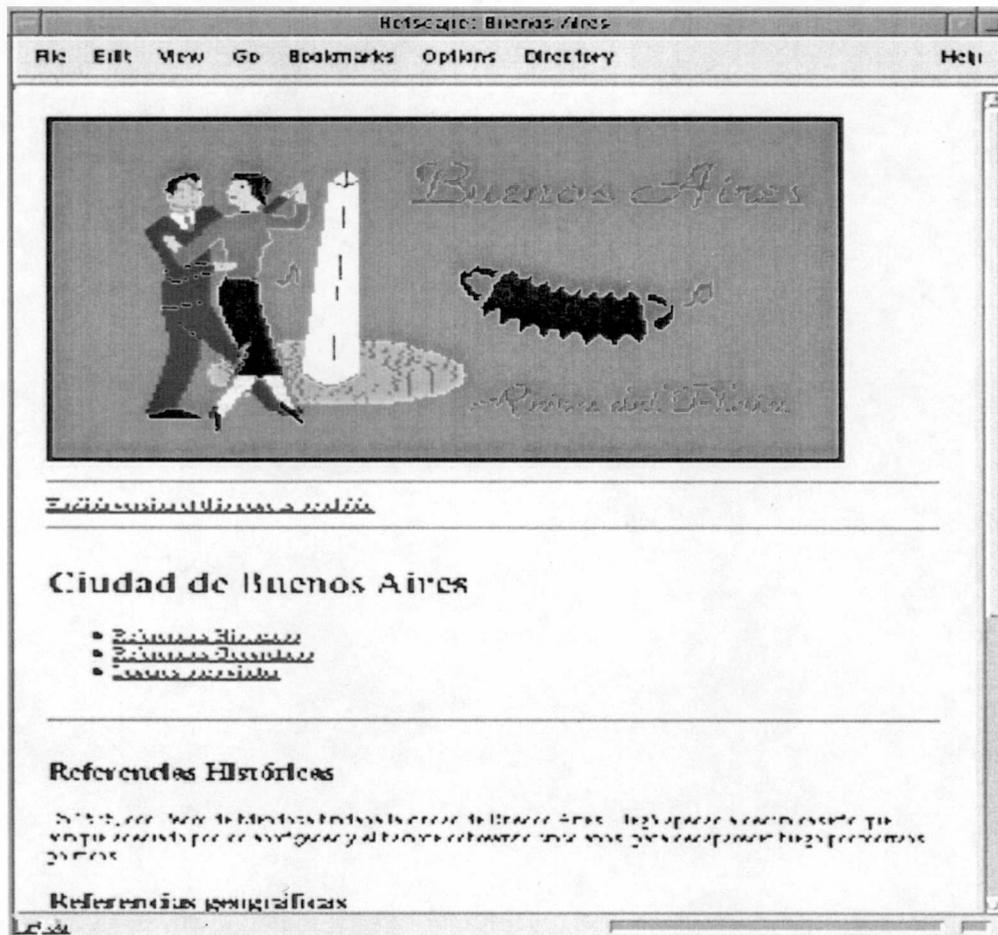
Veamos un ejemplo de un texto HTML (reducido para clarificar el ejemplo) y su respectiva imagen que podemos encontrar en el servidor de WWW www.recyt.net.

```

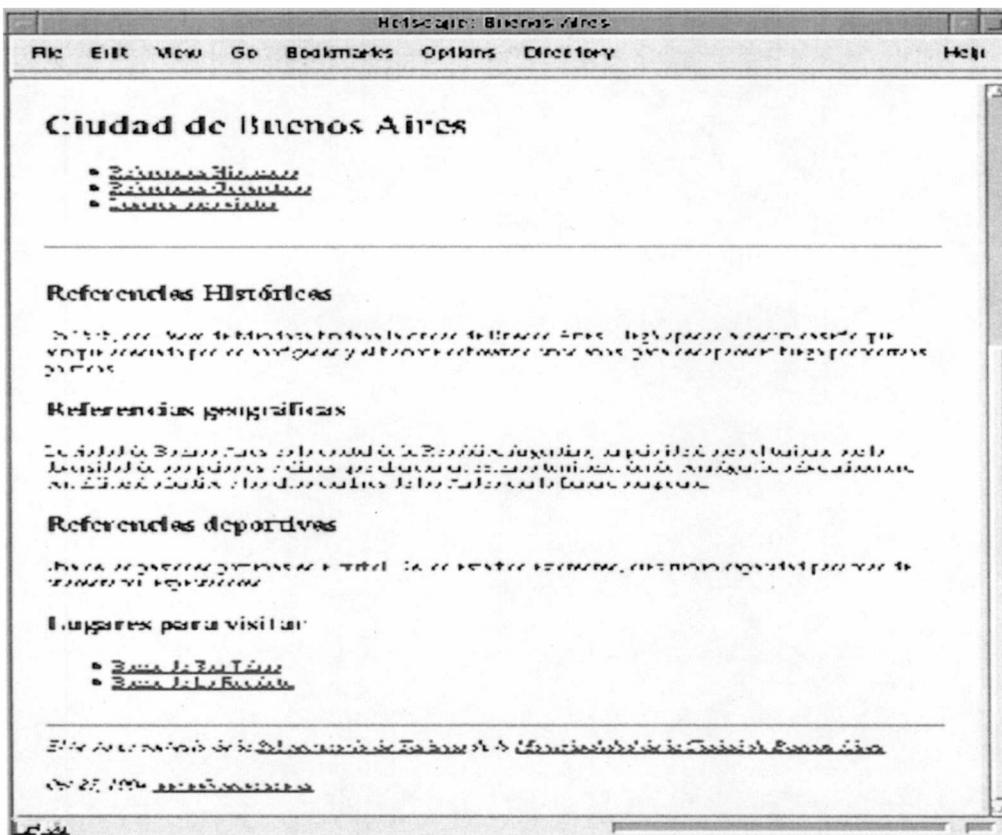
<HEAD>
<TITLE>Buenos Aires</TITLE>
</HEAD>
<BODY>
<IMG SRC="bue.gif" WIDTH=439 HEIGHT=244>
<HR>
<A HREF="buei.html">English version of this page is available</A>
<HR>
<H2>Ciudad de Buenos Aires</H2>
<UL>
<LI><A HREF="#histo">Referencias Hist&oscaracute;ricas</A>
<LI><A HREF="#geo">Referencias Geogr&oscaracute;ficas</A>
<LI><A HREF="#depo">Referencias Deportivas</A>
<LI><A HREF="#visit">Lugares para visitar</A>
</UL>
<HR>
<A NAME="histo"><H3>Referencias Hist&oscaracute;ricas</H3></A>
En 1536, don Pedro de Mendoza fundaba la ciudad de Buenos Aires. Lleg&oscaracute; apenas a ser un caser&oscaracute;o que aunque asediado por los abor&oscaracute;genes y el hambre sobrevivi&oscaracute; cinco a&oscaracute;ntilde;os, para desaparecer luego por motivos pol&oscaracute;ticos.
<P>
<A NAME="geo"><H3>Referencias geogr&oscaracute;ficas</H3></A>
La ciudad de Buenos Aires es la capital de la Rep&oscaracute;blica Argentina, un pa&oscaracute;s ideal para el turismo por la diversidad de sus paisajes y climas que abarcan un extenso territorio, donde comulgan la selva misionera con el litoral atl&oscaracute;ntico y las altas cumbres de los Andes con la llanura pampeana.
<P>
<A NAME="depo"><H3>Referencias deportivas</H3></A>
Una de las pasiones porte&oscaracute;as es el
<A HREF="gopher://gopher.recyt.net/11/infogral/notidepo">f&oscaracute;tbol</A>.
De los estadios existentes, diez tienen capacidad para mas de cuarenta mil espectadores.
<P>
<A NAME="visit"><H3>Lugares para visitar</H3></A>
<UL>
<LI><A HREF="santelmos.html">Barrio de San Telmo</A>
<LI><A HREF="recoletas.html">Barrio de La Recoleta</A>
</UL>
<HR>
<ADDRESS>
El texto es cortes&oscaracute;a de la <A HREF="mcbas.html">Subsecretar&oscaracute;a de Turismo</A> de la <A HREF="mcbas.html">Municipalidad de la Ciudad de Buenos Aires</A>.
<P>
Oct 27, 1994.<A HREF="/msg/dir.html">www@secyt.gov.ar</A>
</ADDRESS>

```

La imagen que vemos por pantalla es la siguiente:



1.



1

GLOSARIO

ANSI:

American National Standards Institute Esta organización es responsable de aprobar los estándares en muchas áreas, incluyendo las áreas de computadoras y comunicaciones. Los estándares que son aprobados por esta organización son llamados generalmente ANSI estandar. ANSI es un miembro de la ISO.[RFC/93]

Anonymous ftp:

Ver ftp anónimo.

APDU - Application Protocol Data Unit:

Es la unidad de datos transferida entre un cliente y un servidor.

Aplicación:

Un programa que realiza una función directamente por el usuario. Los clientes FTP, Mail, y telnet son ejemplos de aplicaciones de red.[RFC/93]

BBS:

Ver Bulletin Board Systems.

Bitnet:

Es una red académica que provee los servicios de correo electrónico y transferencia de archivos utilizando un protocolo de store-and-forward.[RFC/93]

Broadcast:

Es un tipo especial de transmisión en la cual todos los nodos de la red están dispuestos a recibir.[RFC/93]

Bulletin Board Systems :

Es un sistema para almacenamiento de mensajes y archivos, a menudo referente a un tema en particular. Son mantenidos por una única persona o por una organización. Pueden ser accedidos mediante teléfono o si esta en Internet mediante telnet[HAH/94]

Capa:

Las redes de comunicación utilizadas para computadoras puede ser organizadas en un conjunto de protocolos mas o menos independientes, cada uno en capas o niveles diferentes. La capa mas baja maneja directamente las comunicaciones host-a-host entre el hardware de estos. La capa mas alta la constituye las aplicaciones. Cada capa esta construida sobre la otra. Para cada capa, los programas en los diferentes host utilizan el protocolo apropiado para comunicarse con el otro. TCP/IP tiene 5 capas y OSI tiene 7. La ventaja de las distintas capas de protocolo es que el método de pasar información de una a otra esta especificado claramente como parte del conjuntos de protocolos y cambia dentro de la misma capa para prevenir afectar a las otras. Esto simplifica la tarea de diseñar y mantener los programas de comunicación.[RFC/93]

Capa de Aplicación:

Es la capa superior en la pila de la capas de red. Se refiere a la semántica de trabajo. Por ejemplo formato de los mensajes de correo electrónico.[RFC/93]

CCITT:

Comité Internacional de telegrafía y telefonía. Esta organización es parte de la Unión Internacional de telecomunicaciones (ITU) y es responsable de realizar las recomendaciones técnicas sobre telefonía y sistemas de comunicaciones de datos.[RFC/93]

Cliente:

Un proceso o computadora que requiere un servicio de otra computadora o proceso. [RFC/93]

Connection-oriented:

Ver orientada a conexión.

Connectionless:

Ver sin conexión.

Correo Electrónico:

Es un sistema donde un usuario puede intercambiar mensajes con otro usuario o grupos de usuarios a través de una red de comunicaciones. Es uno de los servicios mas populares de Internet.[RFC/93]

CSO - Computing Service Office:

Los servidores CSO pertenecen ala familia de programas llamados directorio de paginas blancas. Utilizados para encontrar usuarios en Internet.[g2]

Dirección internet :

Es una dirección única que identifica a un nodo que esta en internet. Una dirección Internet (I con mayúscula) identifica en forma única a un nodo sobre Internet.[RFC/93]

Dirección IP:

Dirección de 32 bits definida por el protocolo Internet en STD 5, RFC 791. Es usualmente representada con una notación decimal entre puntos.[RFC/93]

DNS - Domain Name Systems:

Es un servicio de consulta de datos, duplicado y distribuido. La función principal es buscar la dirección IP basándose en el nombre del host. El estilo de nombres de host usados actualmente en Internet es "nombre de dominio" ("domain name") pues ellos son utilizados para buscar cualquier cosa en el DNS. Ejemplo:.EDU (educación),.GOV(gobierno), etc. Esta definido en la RFC 1034,1035.[RFC/93]

E-mail:

correo electrónico.

File Transfer:

Ver transferencia de archivos

File Transfer Protocol:

Protocolo de transferencia de archivos.

Finger:

Programa que presenta la información sobre todos o un usuario en particular que están logoneados en un sistema local o remoto. Típicamente muestra el nombre completo, el tiempo inactivo, terminal. También puede mostrar otro tipo de información que desee el usuario.[RFC/93]

FTP - File Transfer Protocol:

Ver protocolo de transferencia de archivos.

Ftp anónimo:

Es una convención para permitir a usuarios de Internet transferir archivos de o hacia maquinas sin restricciones en el acceso.[HAH/94]

Gateways:

Es un programa/dispositivo que pasa datos entre redes que tienen funciones similares pero su implementación es distinta. Un router es un gateways de nivel 3 (capa de red) y un gateways de mail es uno de nivel 7 (capa de aplicación).[RFC/93]

Host:

Se le llama host a una computadora que permite a los usuarios comunicarse con otra sobre una red. los usuarios se comunican a través de programas de aplicación como correo electrónico, telnet, FTP.[RFC/93]

HTML - HiperText Markup Language

Ver lenguaje para señalización de hipertexto.

HTTP - Hypertext Transfer Protocol

Ver protocolo de transferencia de hipertexto.

internet:

Es una red pero el termino "internet" es usualmente usado para referirse a un conjunto de redes interconectadas con ruteadores.[RFC/93]

Internet:

Es una gran internet en el mundo. Tiene 3 niveles de jerarquía compuesta de un "backbone" (ejemplo: NFSNET, MILNET), redes intermedias y redes pequeñas. Es una red multiprotocolo.[RFC/93]

IP:

Protocolo Internet.

ISO - Organización Internacional de Estándares:

Organización responsable de crear los estándares en muchas áreas incluyendo las comunicaciones y computadoras. Sus miembros son pertenecientes a organizaciones de 89 países en la que se incluye ANSI por los EE.UU. Fue fundada en 1946.[RFC/93]

Lenguaje para señalización de hipertexto - HTML:

Lenguaje utilizado para representar instancia de un documento estructurado. Utilizado por WWW.

Listas de correo:

Una lista de correo electrónico es utilizada para enviar mensajes a un grupo de personas. Generalmente es usada para discutir ciertos conjuntos de temas. Puede ser moderada. esto significa que antes de enviarse un mensaje al grupo se envía a un moderados quien determina si enviar o no el mensajes a la lista.[RFC/93]

Listserv:

Ver servidor de listas.

NNTP - Network News Transfer Protocol:

Ver protocolo de transferencia de news.

Orientada a conexión

Método de comunicación de datos en cual el proceso de comunicación pasa a través de 3 fases bien definidas: establecimiento de conexión, transferencia de datos y finalización de la conexión. TCP es un protocolo orientado a conexión.[RFC/93]

OSI-Interconexión de sistemas abiertos:

Es un conjunto de protocolos diseñados por el comité ISO para ser el estandar de arquitectura de redes de computadoras.[RFC/93]

OSI-Modelo de Referencia

Es una estructura de 7 capas diseñadas para describir una arquitectura de red de computadoras y la forma que los datos pasan a través de ellas. este modelo fue desarrollado por la ISO en 1978 para definir en forma clara las interfaces para los vendedores de redes y proveer a los usuarios con una guía conceptual en las construcciones de estas.[RFC/93]

Port:

Es un valor de multiplexación de la capa de transporte. Cada aplicación tiene un único número de port asociado con el.[RFC/93]

Protocolo:

Es una descripción formal de formatos de mensajes y reglas entre dos computadoras que deben intercambiar estos mensajes. Los protocolos pueden ser descriptos a bajo nivel de detalle como por ejemplo cuantos bits o bytes son enviados a través del cable o a alto nivel como por ejemplo la forma en cual dos programas transfieren un archivo a través de Internet.[RFC/93]

Protocolo de Control de Transmisión - TCP:

Es un protocolo estandar de Internet definido en STD 7, RFC 793. Es orientado a conexión.[RFC/93]

Protocolo de Transferencia de Archivos - FTP:

Protocolo que permite a un usuario trabajando en un host acceder y transferir archivos desde o hacia otro host sobre una red. FTP también es el nombre del programa que el usuario invoca para ejecutar el protocolo. Esta definido en le RFC 959, STD 9.[RFC/93]

Protocolo de Transferencia de Hipertexto - HTTP

Protocolo utilizado por WWW para transferencia de hipertexto

Protocolo de Transferencia de mail - SMTP:

Protocolo para transferir mensajes entre computadoras. Es un protocolo de transferencia entre servidores. Se utilizan otros protocolos para acceder a los mensajes. Esta definido en le RFC 821, STD 10.[RFC/93]

Protocolo de Transferencia de News -NNTP:

Protocolo para distribución, recuperación y publicación de artículos de news. Esta definido en la RFC 977.[RFC/93]

Protocolo Internet:

El protocolo Internet, definido en STD 5, RFC 791, es la capa de red para el protocolo TCP/IP. Este es un protocolo sin conexión de transferencia de paquetes.[RFC/93]

RFC:

Es una serie de documentos que describen el conjunto de protocolos de Internet y experimentos relativos. Comenzó en el año 1969. Todos los estandares de Internet son desarrollados basándose en las RFC.[RFC/93]

Red de computadoras:

Una red de computadoras es un sistema de comunicación de datos que interconecta sistemas de computación en diferentes sitios. Una red puede estar compuestas de otras.[RFC/93]

Relevance Feedback:

Método para encontrar información en el cual el resultado de una búsqueda es utilizada como entrada en otra búsqueda. En wais, reelevance feedback permite especificar búsquedas que son difícil de hacerlo en lógica booleana.[KOC/93]

Servidor:

es un proveedor de recursos. Ejemplo: servidores de archivos y servidores de nombres

Servidor de Listas:

Es un sistema de mensajes. Las comunicaciones se basan en el uso del e-mail y le permite a sus usuarios mantener distintos tipo de discusiones como también compartir información. Originalmente fue diseñado para la red Bitnet.[RFC/93]

Sin conexión:

Método de comunicación en cual la comunicación ocurre entre host sin una configuración previa. Los paquetes entre los host pueden tomar diferentes rutas pues cada paquete es independiente del otro. UDP es un protocolo sin conexión.[RFC/93]

SMTP - Simple Mail Transfer Protocol:

Ver Protocolo de transferencia de mail

STD:

Es un subconjuntos de una serie de estandares que especifican los estandares de Internet. La lista de estos estandares están descriptas en el STD 1.[RFC/93]

TCP - Transmission Control Protocol:

Ver protocolo de control de transmisión.

TCP/IP:

Protocolo de control de transmisión sobre protocolo de Internet. Es una forma común de referenciarse al conjunto de protocolos de transporte y aplicación que corren sobre IP.[RFC/93]

Telnet:

Es un protocolo estandar de Internet para conectar a un usuario en una terminal remota. esta definido en STD 8, RFC 854 y extendidas con opciones por muchas otras RFC.[RFC/93]

Transferencia de archivos:

Copia de un archivo de una computadora a otra sobre una red de computadoras.[RFC/93]

Usenet:

El termino se utiliza para varias cosas simultáneamente:

Es una colección de miles de temas organizados en grupos (newsgroup) los cuales cubren desde supercomputación, cocina etc. Su distribución se realiza desde el mundo entero hasta una simple institución.

Se denomina Usenet a la red de computadoras que reciben los grupos de news. [KOC/93]

X400:

Son los estandar para correo electrónico de la CCITT e ISO.[RFC/93]

X500:

Es un estandar CCITT e ISO para servicio de directorio,[KOC/93]

BIBLIOGRAFIA

[ALB/92] Bob Alberti, Farhad Anklesaria, Paul Linder, Mark Mchill, Daniel Torrey:

The Internet Gopher Protocol - The distributed documents Search and Retrieval Protocol, Universidad de Minnesota Microcomputer and Workstation Network Center, 1992.

[BER/92] Tim Bernes-Lee, Robert Caillau:

"World Wide Web". Paper que se presentó en la conferencia de Computing in High Energy Physic, del 23-27 de septiembre de 1992, Francia.

[CAI/92] Tim Bernes-Lee, Robert Cailliau, Jean-Françoise Groff, Bernd Pollermann:

"World Wide Web: The Information Universe ", publicado en Electronic Networking:research, Applications and Policy, Vol2, No. 1, 1992

[CAI/93] Tim Bernes-Lee

"The World Wide Web Iniciative", Inet'93 , Proceedings Agosto 1993.

[DAN] Peter Danzing, Katia Obraczka and Shih-Hao Li:

Internet Resource Discovery Services, Computer Science Departament University of Southern California

[DAV/90] Franklin Davis, Brewster Kahle, Harry Morris, Jim Salem, Tracy Shen, Rod Wang, Jhon Sui, Mark Grinbaum:

WAIS Interface Protocol Prototype Functional Specification, Thinking Machine Corporation, Dow Jones & Company, Inc, 1990

[FAL/94] Bennett Falk :

"The Internet Roadmap", Sybex, 1994, pags 178-199

[GIL/94] Paul Gilster:

Finding it on the Internet, John Wiley & Sons Inc., 1994

[GOP/92] Internet Gopher Server Documentation - University of Minnesota. Documentación que acompaña al servidor Gopher, v.1.1, 1992.

[HAH/94] Harley Hahn & Rick Stout:

The Internet Complete Reference, MacGrawHill.,1994,

[KAH/89] Brewster Kahle:

Wide Area Information Server Concepts, Thinking Machine Corporation

[KEH/94] Brenda Kehoe:

Zen and the Art of Internet A beginner's Guide, Prince Hall, 3ra. edición,1994.

[KOC/93] Jonathan Kochmer and NorthWestNet:

Internet Passport, NorthWestNet, 1993.

[KRO/94] Ed Krol,

The Whole Internet user Guide & Catalog, O'Reilly & Associates, 1994, 2da. edición.

[LAQ/93] Tracy Laquey with Jeanne Ryer:

The Internet Companion, Addison Wesley Publishing Company,1993

[LEE/92] Tim Bernes-Lee,

"Word Wide Web Book" , 1992

[LEE/93] Tim Bernes-Lee, Daniel Connolly :

Hypertext Markup Languaje: A representation of textual

**Information and Metainformation for Retrieval and Interchange,
13 de Julio de 1993.**

**[LIU/94] Cricket Liu, Jerry Peek, Russ Jones, Brian Buus & Adrian Nye:
Managing INTERNET Information Services, O'Reilly & Associates
1994, primera edición.**

**[MAR/92] April Marine
Internet: Getting Started, Internet Information Series, 1992,
pag194**

**[MOR/94] Eric Lease Morgan
Wais & Gopher Servers A guide for Internet EndUsers, Meck
Lermendia, 1994.**

**[OTT/94] Peter Otter
The Information Superhighway Beyond the Internet, Que, 1994.**

**[RFC/93] User Glossary Working Group:
RFC 1392, Internet Glossary, 1993**

**[SCH/92] Michael Schwartz, Alan Emtage, Brewster Kahle, Clifford
Neuman:
A Comparison of Internet Resource Discovery Approaches,
Agosto 1992**

[Z39/91] Ansi Z39.50 Version 2 THIRD DRAFT, Mayo 1991



BIBLIOTECA
FAC. DE INFORMÁTICA
U.N.L.P.

INDICE

A

<i>Anonymous ftp</i>	108
<i>ANSI</i>	108
<i>APDU</i>	108
<i>Aplicación</i>	108
<i>Application Protocol Data Unit</i>	108
<i>Archie</i>	15
<i>Arquitectura</i>	18
<i>Como acceder</i>	20
<i>Como Trabaja</i>	16
<i>Vía correo electrónico</i>	24
<i>Vía telnet</i>	20
<i>Vía un cliente</i>	23
<i>Whatis</i>	22

B

<i>BBS</i>	108
<i>BIBLIOGRAFIA</i>	117
<i>Bitnet</i>	108
<i>Broadcast</i>	108
<i>Bulletin Board Systems</i>	109

C

<i>Capa</i>	109
<i>Capa de Aplicación</i>	109
<i>CCITT</i>	109
<i>Cliente</i>	109
<i>Cliente/servidor-Modelo</i>	2
<i>Computing Service Office</i>	110
<i>Connectionless</i>	110
<i>Connection-oriented</i>	110
<i>Consideraciones Finales</i>	73
<i>Correo Electrónico</i>	110
<i>CSO</i>	39, 110

D

<i>DGC</i>	18
<i>Dirección internet</i>	110
<i>Dirección IP</i>	110
<i>DMC</i>	18
<i>DNS</i>	110
<i>Domain Name Systems</i>	110

E

<i>E-mail</i>	111
---------------------	-----

F

<i>File Transfer</i>	111
<i>File Transfer Protocol</i>	111
<i>File Transfer Protocol:</i>	111

Sistemas de Hallazgo y Recuperación de Recursos en Internet 121

<i>Finger</i>	111
<i>FTP</i>	111, 114
<i>Ftp anónimo</i>	111
G	
<i>Gateway</i>	111
GLOSARIO	108
<i>Gopher</i>	26
<i>Arquitectura</i>	32
<i>Como acceder</i>	42
<i>Interface</i>	31
<i>Jughead</i>	45
<i>Modelo de los Datos</i>	27
<i>Protocolo Gopher</i>	33
<i>Protocolo Gopher+</i>	41
<i>Servidor Gopher</i>	42
<i>Veronica</i>	44
<i>Vía telnet</i>	44
<i>Vía un cliente</i>	42
H	
<i>HiperText Markup Languaje</i>	111
<i>Hipertext Transfer Protocol</i>	112
<i>Host</i>	111
<i>HTML</i>	111
<i>HTTP</i>	112, 114
I	
<i>internet</i>	112
<i>IP 112</i>	
<i>ISO</i>	112
J	
<i>Jughead</i>	45
K	
<i>KIS (Knowbot Information Server)</i>	5
L	
<i>Lenguaje para señalización de hipertexto</i>	112
<i>Listas de correo</i>	112
<i>Listserv</i>	112
N	
<i>Netfind</i>	9
<i>Via cliente</i>	9
<i>Via telnet</i>	9
<i>Network News Transfer Protoco</i>	112
<i>NNTP</i>	112, 114
O	
<i>Organización Internacional de Estándares</i>	112
<i>Orientada a conexión</i>	113

Sistemas de Hallazgo y Recuperación de Recursos en Internet 122

<i>OSI-Interconexión de sistemas abiertos</i>	113
<i>OSI-Modelo de Referencia</i>	113
P	
<i>Port</i>	113
<i>Prospero</i>	13
<i>Protocolo</i>	113
<i>Protocolo de Control de Transmisión</i>	113
<i>Protocolo de Transferencia de Archivos</i>	114
<i>Protocolo de Transferencia de Hipertexto</i>	114
<i>Protocolo de Transferencia de mail</i>	114
<i>Protocolo de Transferencia de News</i>	114
<i>Protocolo Internet</i>	114
R	
<i>Red de computadoras</i>	114
<i>Relevance Feedback</i>	115
<i>RFC</i>	114
S	
<i>Servidor</i>	115
<i>Servidor de Listas</i>	115
<i>Simple Mail Transfer Protocol</i>	115
<i>Sin conexión</i>	115
<i>Sistemas de Menus y Web</i>	71
<i>SMTP</i>	114, 115
<i>STD</i>	115
T	
<i>TCP</i>	113, 115
<i>TCP/IP</i>	115
<i>Telnet</i>	116
<i>Transferencia de archivos</i>	116
<i>Transmission Control Protocol</i>	115
U	
<i>UAC</i>	18
<i>Usenet</i>	116
V	
<i>Veronica</i>	44
W	
<i>WAIS</i>	47
<i>Arquitectura</i>	47
<i>Características</i>	55
<i>Como Acceder</i>	51
<i>Como realizar una búsqueda</i>	52
<i>Como Trabaja</i>	50
<i>Protocolo Z39.50</i>	49
<i>Relevance Feedback</i>	55
<i>Selección de los documentos mas adecuados</i>	56
<i>Via Gopher, WWW</i>	51

Sistemas de Hallazgo y Recuperación de Recursos en Internet 123

<i>via Telnet</i>	51
<i>Via un cliente</i>	51
<i>Wais y Web</i>	71
WHOIS	1
<i>Via cliente</i>	2
<i>Via correo electrónico</i>	4
<i>Via telnet</i>	4
WWW	58
<i>Arquitectura</i>	62
<i>Como acceder</i>	65
Direccionamiento	
<i>Gopher</i>	88
<i>HTTP</i>	86
<i>para servidores WAIS</i>	89
<i>Telnet</i>	87
<i>Evitando caracteres ilegales</i>	92
<i>Hipertexto, ¿Qué es?</i>	59
<i>Introducción</i>	58
<i>Lenguaje HTML</i>	94
<i>modelo de datos</i>	60
<i>Protocolos</i>	63
<i>Relaciones entre WWW y los demás sistemas</i>	71
<i>Servidores</i>	70
<i>Sintaxis de direccionamiento WWW: BNF</i>	90
<i>Técnicas de Navegación y Herramientas</i>	67
<i>URL-Universal Resource Locator</i>	60
<i>Vía Telnet</i>	66
<i>Vía un cliente</i>	65
<i>vía una interface gráfica</i>	67
<i>vista del proveedor de información</i>	68
 X	
<i>X400</i>	116
<i>X500</i>	116

DONACION.....

TES
9414

\$.....

Fecha..... 1-3-06

Inv. E..... Inv. B. 2505



BIBLIOTECA
FAC. DE INFORMÁTICA
U.N.L.P.