White, Bob

1) DEFINITION SERVER & BROWSER

browser: like MIDAS, XMOSAIC, LYNX, MACMOSAIC, linemode

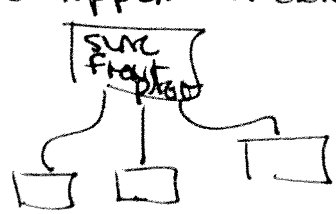
* permits the display of hypertext docs (i.e. docs that contain links)
* display of other files (plain/text)
* links ~~to~~ files to other ~~services~~ APPLICATIONS, e.g. GHOSTVIEW, xv, etc. (done via file extension)
* connects to other Internet services (special protocols)
* communicates with HTTP servers.

SERVERS. * provide docs to browsers (daemons)
* ~~RUN SPECIAL~~ capture output of special processes.

* [don't have to have a server in order to use a browser]

2) PAGES: A hypertext doc. (i.e. contains links)

enter W3 @ some page; often the home page or root/entry page (but this doesn't mean hierarchical); succession of pages @ a site often appear hierarchical i.e. } hence **WEB**



but don't have to be.

technically can enter @ any point

3) Responsibility of pages will "trickle down" to the group or individual level.
(space is there — look @ SLD, BES, other)

just need to define information.

(4) what kind of info do we mean?
A) files, documents, pictures, etc.
B) Access to various network services (ftp, gopher, news...)
C) links to [illegible] W3 servers (e.g. SLAC, CERN, Fermi) which can
1) serve documents
2) provide index searching (e.g. SPIRES, VM HELP)
3) capture output of specially designed server processes.
D) whatever is of interest to your users and/or group

look @ Networking page

(5) So the 1st step is to write a hypertext page:
* written in a markup language called HTML (Hypertext Markup Language)

* tags in HTML
→ describe document formatting (headers, titles, lists, bullets, columnization)
→ provide syntax for include of graphics
→ provide a syntax for defining links to other documents. links are highlighted in browsers

[Examples]

look @ source of Networking page

→ Author of page is usually @ bottom

{ UNIVERSAL
{ UNIFORM
{ UNIQUE

(6) URLs — define complete information necessary for a link

~~the~~ FORM of URL:

* used to HTML documents in conjunction with anchor defs. <a . . .
  ~~&~~ and ~~HREF~~ HREFs (hypertext Reference)

* May also be used as a full document definition (such as open document or GO)

~~* full or absolute URLs~~
~~* partial or relative URLs~~

Also an increasing trend amongst W3 users to use URLs as References (as in bibliographies)

[For ex. give URLs for HTML Reference docs or W3 Reference

* Full or absolute URLs
* partial or relative URLs (local)

→ Full usually implies the use of some server (i.e. http, gopher, ftp, etc.)
→ partial can be used to move between documents in shared directories or networks.

@SLAC.
→ so theoretically all Unix references can be – due to inf: partial

[EXAMPLES]

(7) Servers are used to
* serve up documents between platforms (i.e. Unix files to VM) or when absolute URLs are used
* provide some security
→ files/directories can be blocked out
→ aliases can hide directory paths (by mapping)
→ sites can be blocked out according to the domain of browser requests (e.g. Prodtrak only visible @ SLAC)
* special processes.
The next step can be writing these — capture direct output ~~socket~~ or creates a hypertext document. and sends thru socket.

(8) Using your pages —

→ WWW_HOME defines <u>root/entry</u> page now ~~set~~ set to SLAC home page

→ invoke browser with file name of page

→ set aliases

→ <u>IF</u> you want your page worked into the SLAC hierarchy, contact me or Joan Winters

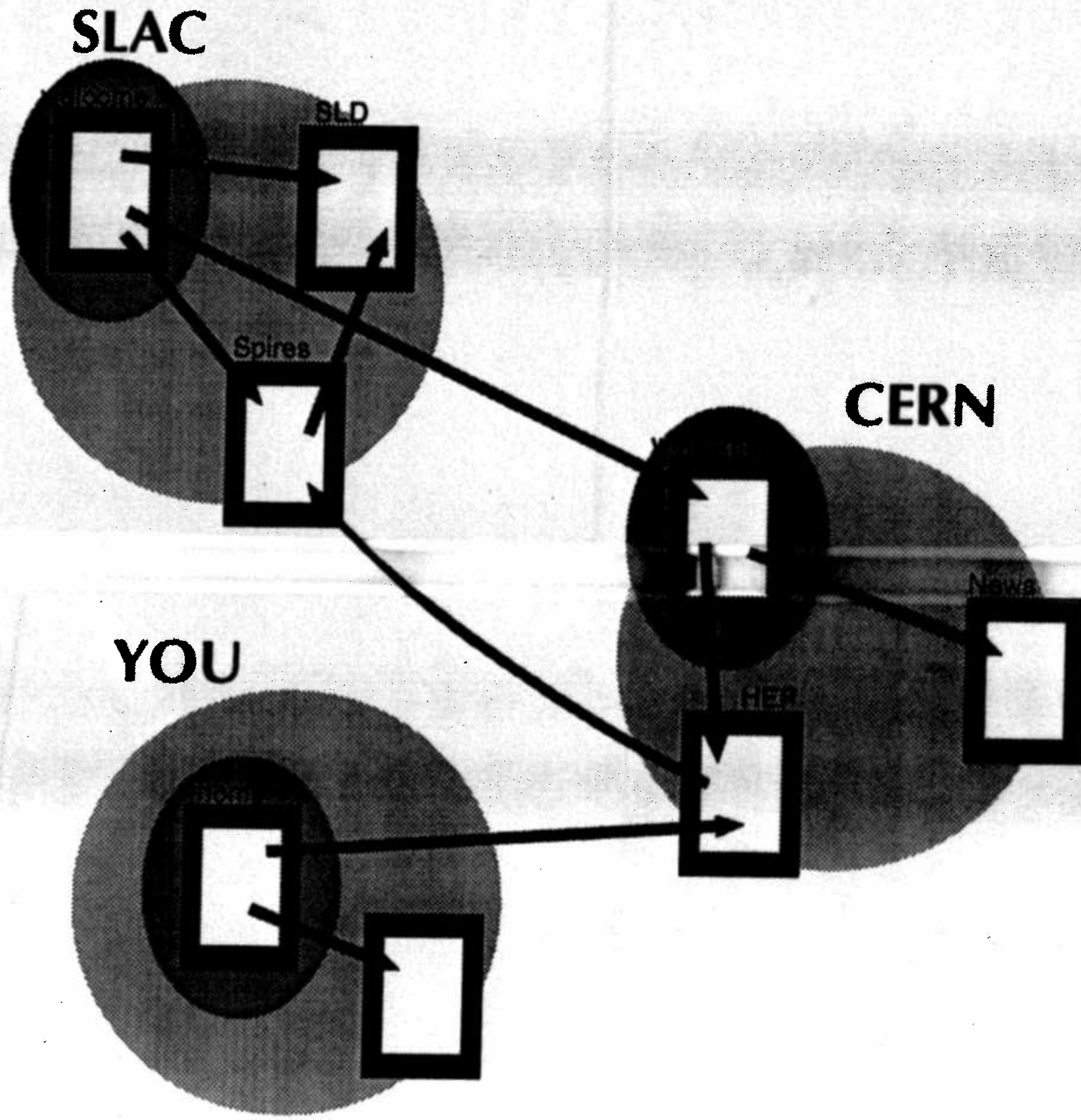
What is WWW?

- WorldWideWeb (WWW or W3) is a network distributed information organization system. Using hypertext links, WWW permits document retrieval and keyword indexed database searches. WWW also provides gateways to other information retrieval systems.

How Does WWW Look?

- The usual WWW client interface (browser) is designed to resemble a menu (page) with active buttons or keys.
- Buttons (links) on the page may be on many different machines anywhere on the Internet. As the user pushes buttons and browses one file after another, he/she goes from one machine to another, exploring a huge information network (web).
- Buttons/links need not be sequential/hierarchical (i.e., hypertext metaphor).

An actual situation



Command Quick-Reference

a href="*protocol://location/file#destination*"
/a .... Start/stop link anchor

a name="*anchorname*"
/a .... Start/stop destination anchor

title
/title .... Start/stop WWW browser title

h1
/h1 .... Start/stop a level 1 heading

p .... Paragraph break

h2
/h2 .... Start/stop a level 2 heading

i
/i .... Start/stop italics

b
/b .... Start/stop boldface

code
/code .... Start/stop fixed width font

<pre>
</pre> .... Start/stop pre-formatted text

img *align=position* **src="*pictureURL*"**
.... Insert a graphic (`.GIF/.XBM`)

ol .... Start ordinal list
li .... Specify a list item
/ol .... Stop ordinal list

ul .... Start bulleted list
li .... Specify a list item
/ul .... Stop bulleted list

dl .... Start description list
dt .... Description title
dd .... Description text
/dl .... Stop description list

Universal Resource Locator

http		
file		
gopher	***: // host : port /path / path***	?search #anchor
wais		
news		

Examples:

- file://info.cern.ch/pub/www/doc/url1.ps
- http://crnvmc.cern.ch:2784/FIND?sgml+tag
- fred.html#par1
- #anc
- news:comp.windows.x

Reference to HTML Primer:

http://www.ncsa.uiuc.edu/demoweb/html-primer.html

Reference to URL Primer:

http://www.ncsa.uiuc.edu/demoweb/url-primer.html

Reference to HTML Specifications:

ftp://info.cern.ch/pub/www/doc/html-spec.ps

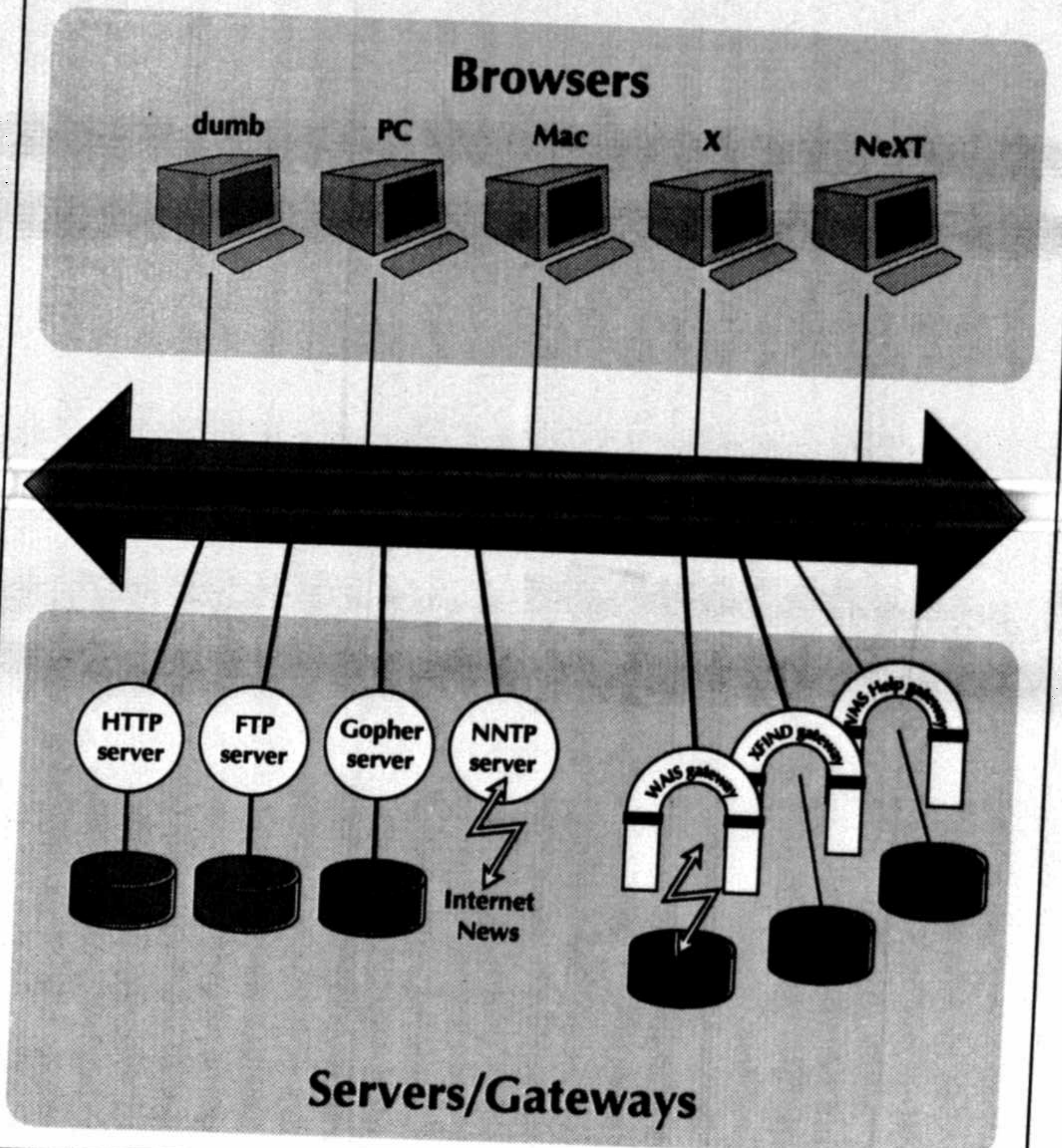
Reference to an On-Line WWW Tutorial:

http://info.cern.ch/hypertext/WWW/Talks/General.html

Reference to the WWW FAQ:

ftp://rtfm.mit.edu/pub/usenet/news.answers/www/faq

Architecture



Universal Resource Locator

http		
file		
gopher	***: // host : port /path / path***	?search #anchor
wais		
news		

Examples:

- file://info.cern.ch/pub/www/doc/url1.ps
- http://crnvmc.cern.ch:2784/FIND?sgml+tag
- fred.html#par1
- #anc
- news:comp.windows.x

Protocols

A number of existing protocols, and one new protocol, form the set with which a WWW browser is equipped. Standard protocols used are:

- FTP (File Transfer Protocol)
- NNTP (Network News Transfer Protocol)

A WWW browser is also equipped to handle the protocols of two other information retrieval systems:

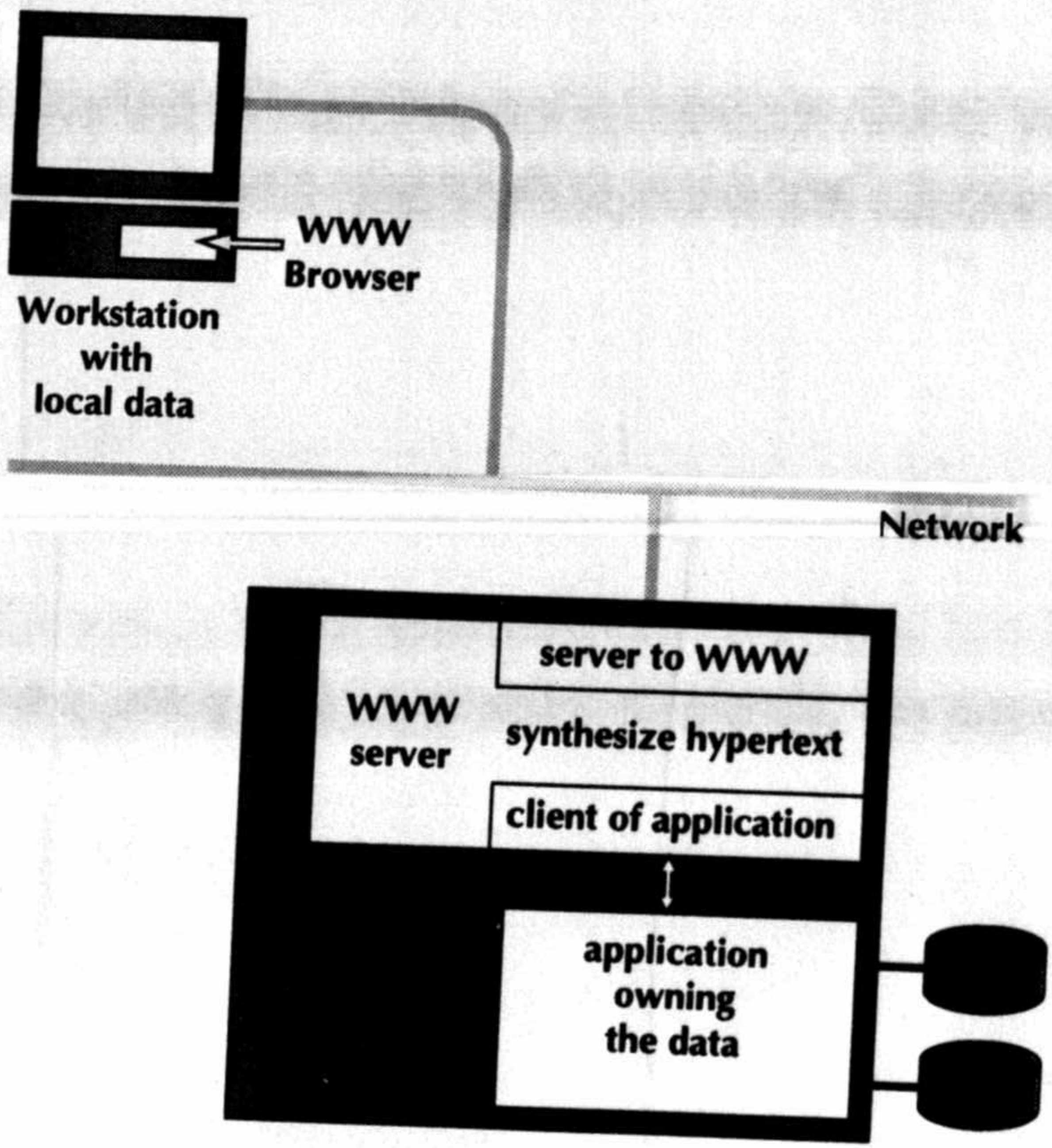
- WAIS (Wide Area Information Servers)
- Gopher

A new protocol, HTTP (Hypertext Transport Protocol), was developed to allow document retrieval and index search.

```
<title>Networking</title>
<h1>SLAC Computer Services (SCS) Networking</h1>
<h2>General Information (non-SLAC specific) including Frequently Asked Questions (FAQs):</h2>
<ul>
<li> Ethernet <a HREF=ftp://dorm.rutgers.edu/pub/novell/DOCS>FAQs</a> and
     <a HREF= ftp://ftp.lcs.mit.edu/pub/map/EtherNet-codes> Manufacturer Codes</a>
<li> <a HREF=ftp://ftp.nwnet.net/user-docs/cell-relay/archive/>ATM Archives</a>
<li> <a HREF=ftp://rtfm.mit.edu/pub/usenet/news.answers/x-faq>X Windows System FAQs</a>
<li> Packetized Video Conferencing <a HREF=ftp://venera.isi.edu/mbone/faq.txt>Multicast Backbon
e (MBONE) FAQs</a>, Cornell's Macintosh
     <a HREF=ftp://gated.cornell.edu/pub/video/>CU-SeeMe</a> and
     <a HREF=ftp://ftp.es.net/pub/mailing-lists/mail-archive>Mailing list Archives</a>
<li> Energy Research Video Network (ERVN) <a HREF=ftp://sscvx1.ssc.gov/network/video_final.ps>p
roposal</a>,
     <a HREF=ftp://ftp.es.net/documents/videoconf.ps> what to buy</a>,
     <a HREF=ftp://nic.hep.net/misc/sept28-ervn-list.ps> contacts</a> and
     <a HREF=ftp://ftp.es.net/misc/ervn-net.ps> map</a>
<li> China (PRC) <a HREF= ftp://cnd.org/pub/cinet/china/net-china.faq>networking FAQs</a> and
     <a HREF=ftp://cnd.org/pub/> Archive</a>
<li><a HREF= ftp://ftp.es.net/pub/networking-info/earn/nettools.txt>Network Information Retriev
al Tools (Gopher, WWW ...)</a>
<li><a HREF=gopher://gopher.hep.net:70/11/info_center/simple-times>Simple Times SNMP Newsletter
</a>

</ul>
<p>
<h2> Some Network Information Servers:</h2>
<ul>
<li><a HREF= http://www.hep.net/>HEP Information Server</a>
<li>ESnet Information Server:<a HREF= ftp://ftp.es.net/>FTP</a> access,
    <a HREF=gopher://gopher.es.net:70/> Gopher</a> access and ESnet
    <a HREF=ftp://ftp.es.net/pub/esnet-stats/> usage statistics</a>
<li><a HREF= http://nearnet.gnn.com/GNN-ORA.html>Global Network Navigator</a>
<li><a HREF= ftp://stis.nsf.gov/>NSF Science & Technology Information System</a>
<li> Internet Network Information Centers (NICs) at: <a HREF= ftp://nic.merit.edu/> Merit</a> a
nd the
    <a HREF=ftp://ds.internic.net/pub/internic-info> InterNIC Directory and Database Services</
a>
<li><a HREF=gopher://gopher.lbl.gov:70/11/>LBL Gopher server</a>
<li><a HREF=gopher://calypso.oit.unc.edu:70/11/NII.d>National Information Infrastructure</a>
<li><a HREF=gopher://sunsite.unc.edu:70/11/>Sun Microsystems anonymous ftp archives</a> and
    <a HREF= ftp://sunsite.unc.edu/pub/sun-info/sunergy>Sunergy broadcasts</a><p>
<li><a HREF= ftp://gatekeeper.dec.com/pub/>DEC</a> and
    <a HREF= ftp://ibminet.awdpa.ibm.com/pub/>IBM</a> Information Servers<p>
<li><a HREF= ftp://info.cren.net/>BITNET/CREN Information Server</a><p>
<li><a HREF= ftp://sparkyfs.erg.sri.com/BayLISA/>Large Installed Systems Administration for San
 Francisco Area - part of USENIX</a>
<li><a HREF=gopher://gopher.es.net:70/11/gopher-stuff/X.500>X.500 White Pages (via ESnet Gatewa
y)</a>
</ul>
<h2> SLAC Specific Networking Information:</h2>
<ul>
<li><a HREF=gopher://jupiter.slac.stanford.edu:5070/00/SLAC%20Servers'%20Status>SLAC Servers' S
tatus</a>
<li><a HREF= http://slacvm.slac.stanford.edu:5080/FIND/xwhere.html>SLAC WHEREIS</a>
<li><a HREF=gopher://jupiter.slac.stanford.edu:5070/11/CANDO%20Reports>SLAC Computer And Networ
k Database in Oracle (CANDO) Reports</a>
<li><a HREF=gopher://jupiter.slac.stanford.edu:5070/11/SLAC%20SCS%20Networking//report>SLAC Netw
ork Performance Reports</a>
<li>SLAC Networking <a HREF=gopher://gopher.slac.stanford.edu:5070/11/SLAC%20SCS%20Networking/d
oc>Presentations</a>,
                    <a HREF=ART.HTML>Transparencies</a> and
                    <a HREF=NETFOLKS.HTML>Personnel</a>
</ul>
<h2>Netnews Groups:</h2>
<ul>
<li><a HREF=news:slac.networks>slac.networks</a>
</ul>
</dl>
<address><a HREF=http://slacvm/FIND/binlist?find+name+cottrell>Cottrell</a></address>
```

Easy for providers !



Making available existing data

```
<title>Comparing WWW and Gopher</title>
<h1>Comparing WWW and Gopher</h1>
<h2>Information Retrieval in HEP</h2>
<ul>
<li>
 The promise of easy access to large amounts
of archival data has long been of interest to
physicists.  The continued development of
high-speed international networks,
relatively inexpensive mass-storage units,
and fast, low cost computers have made
such access possible.
<li>
 Examples -
<ul>
<li>
 QSPIRES
<li>
 Journal and document preprints
<li>
 News
<li>
 Whois-type information
</ul>
</ul>


<h2>What is Gopher?</h2>
<ul>
<li> Internet Gopher is a distributed server
document search and retrieval system.
Gopher combines the features of both
electronic bulletin board services (a
hierarchical organization of items) and of
full-text searchable databases (searches based
on the content of documents where all words
in the document are considered keywords).
</ul>

<h2>What is WWW?</h2>
<ul>
<li> WorldWideWeb (WWW or W3) is a network
distributed information organization
system. Using hypertext links, WWW
permits document retrieval and keyword
indexed database searches. WWW also
provides gateways to other information
retrieval systems.
</ul>


<h2>What Gopher & WWW Have in Common</h2>
<ul>
<li> Both the WWW and  Gopher information
retrieval systems are based on a
client/server architecture so that users on a
heterogeneous mix of desktop systems can
```

```
browse, search, and retrieve documents
residing on multiple distributed server
machines anywhere on the Internet.
<li>
 Both WWW and Gopher attempt to make
information sources transparent to the
user.
<li>
 Both WWW and Gopher automatically
handle resource protocols.
</ul>

<h2>How Does Gopher Look?</h2>
<ul>
<li> The Gopher protocol and the usual Gopher
client interface is designed to resemble a
hierarchical file system since a hierarchy is
a good model for locating documents and
services.
<li>
 Therefore,  what the user sees is a "virtual"
directory; items in the listing may be on
many different machines anywhere on the
Internet.  As the user clicks from one
directory to another and browses one file
after another, he/she goes from one
machine to another, exploring a huge
networked file system.
<li>
 File system metaphor implies sequential
resource reference.
</ul>

<h2>How Does WWW Look?</h2>
<ul>
<li> The usual WWW client interface (browser)
is designed to resemble a menu (page) with
active buttons or keys.
<li>
 Buttons (links) on the page may be on many
different machines anywhere on the
Internet.  As the user pushes buttons and
browses one file after another, he/she goes
from one machine to another, exploring a
huge information network (web).
<li>
 Buttons/links need not be
sequential/hierarchical (i.e., hypertext
metaphor).
</ul>

<h2>Comparing Gopher and WWW</h2>
<ul>
<li>
 Links-
<ul>
<li> W3 is based on hypertext documents,
```

and is structured by links between
pages of hypertext. There are no rules
about which documents can point
where.

Gopher is based on individual resources
and servers. It doesn't have the concept
of a link.

Interfaces-

 With Gopher the interface changes
depending upon the resource being
used.

WWW provides a uniform interface to
services. The hypertext concept allows
user to perform two operations - follow
a link and perform a search.

WWW restructures services as hypertext.
Gopher provides services as provided
(usually just there).

Available Resources -

 For both WWW and Gopher, the lowest
common denominator for data is text.

Gopher allows users can also view
graphics files, hear sound files or
retrieve software for their desktop
machines in a variety of formats
(binary, Macintosh BinHex, uuencode).
"Format negotiation" is a future
objective for WWW or a characteristic of
WWW browsers.

Gopher can access services built on
other client/server protocols such as
WAIS servers or Archie servers
transparently; Anonymous FTP servers
look like directories to the Gopher user.

Gopher cannot access WWW services
since it doesn't know how to deal with
hypertext.

WWW can access all Gopher servers.

Expandability -

 Both Gopher and WWW are readily
expandable for local information

```
sources.
<li>
Aapplication developers find it easy to
write and debug Gopher clients and
servers; users typically need no training
in using client software.
<li>
Wwriting WWW servers requires some
knowledge of the WWW  code;
knowledge of TCP/IP helpful; bulk of
server code writing involves hypertext
processing; users would typically need
no training in using client software.
</ul>
<li>
 Support -
<ul>
<li> Gopher support comes from the
University of Minnesota. Code is
available via Anonymous FTP for a wide
variety of platforms.
<li>
 WWW is becoming widely used and
supported in the HEP community -
CERN, SLAC, DESY, IN2P3, NIKHEF,
FermiLab, SSC, et. al. It is now
distributed as a part of CERNLIB. Code is
also available via Anonymous FTP from
info.cern.ch for a wide variety of
platforms.
</ul>
<li>
 Present HEP usage -
<ul>
<li> One of WWW's original design goals was
to provide ready access to
SPIRES/QSPIRES. The WWW developers
are very interested in the future of
SPIRES and will accomodate the needs of
SLAC.
</ul>
</ul>
<address>Bebo White, bebo@slac.stanford.edu</address>
```

Information Retrieval in HEP

- The promise of easy access to large amounts of archival data has long been of interest to physicists. The continued development of high-speed international networks, relatively inexpensive mass-storage units, and fast, low cost computers have made such access possible.

- Examples -

  - QSPIRES

  - Journal and document preprints

  - News

  - Whois-type information

What is Gopher?

- Internet Gopher is a distributed server document search and retrieval system. Gopher combines the features of both electronic bulletin board services (a hierarchical organization of items) and of full-text searchable databases (searches based on the content of documents where all words in the document are considered keywords).

What Gopher & WWW Have in Common

- Both the WWW and Gopher information retrieval systems are based on a client/server architecture so that users on a heterogeneous mix of desktop systems can browse, search, and retrieve documents residing on multiple distributed server machines anywhere on the Internet.

- Both WWW and Gopher attempt to make information sources transparent to the user.

- Both WWW and Gopher automatically handle resource protocols.

How Does Gopher Look?

- The Gopher protocol and the usual Gopher client interface is designed to resemble a hierarchical file system since a hierarchy is a good model for locating documents and services.

- Therefore, what the user sees is a "virtual" directory; items in the listing may be on many different machines anywhere on the Internet. As the user clicks from one directory to another and browses one file after another, he/she goes from one machine to another, exploring a huge networked file system.

- File system metaphor implies sequential resource reference.

Comparing Gopher and WWW

- Links-
    - W3 is based on hypertext documents, and is structured by links between pages of hypertext. There are no rules about which documents can point where.
    - Gopher is based on individual resources and servers. It doesn't have the concept of a link.
- Interfaces-
    - With Gopher the interface changes depending upon the resource being used.
    - WWW provides a uniform interface to services. The hypertext concept allows user to perform two operations - follow a link and perform a search.
    - WWW restructures services as hypertext. Gopher provides services as provided (usually just there).

- Utility -
  - For both WWW and Gopher, the lowest common denominator for data is text.
  - Gopher allows users can also view graphics files, hear sound files or retrieve software for their desktop machines in a variety of formats (binary, Macintosh BinHex, uuencode).
  - "Format negotiation" is a future objective for WWW or a characteristic of WWW browsers.
  - In WWW there is no search capability for non-hypertext information.
  - In Gopher, users can ask index servers to return a list of all documents that contain one or more words. Since an index server does full-text searching, every word in every document is a keyword.

- Available Resources -
    - Gopher can access services built on other client/server protocols such as WAIS servers or Archie servers transparently; Anonymous FTP servers look like directories to the Gopher user.
    - Gopher cannot access WWW services since it doesn't know how to deal with hypertext.
    - WWW can access all Gopher servers.

- Expandability -
    - Both Gopher and WWW are readily expandable for local information sources.
    - application developers find it easy to write and debug Gopher clients and servers; users typically need no training in using client software.
    - writing WWW servers requires some knowledge of the WWW code; knowledge of TCP/IP helpful; bulk of server code writing involves hypertext processing; users would typically need no training in using client software.

- Support -
  - Gopher support comes from the University of Minnesota. Code is available via Anonymous FTP for a wide variety of platforms.
  - WWW is becoming widely used and supported in the HEP community - CERN, SLAC, DESY, IN2P3, NIKHEF, FermiLab, SSC, et. al. It is now distributed as a part of CERNLIB. Code is also available via Anonymous FTP from info.cern.ch for a wide variety of platforms.
- Present HEP usage -
  - One of WWW's original design goals was to provide ready access to SPIRES/QSPIRES. The WWW developers are very interested in the future of SPIRES and will accommodate the needs of SLAC.

What is WAIS?

- WAIS is a search and retrieval system originally designed by Thinking Machines, Inc. for "free text" searching.
- WAIS attempts to remove the need for direct correspondence between searchable data and retrievable objects.
- The server provides a search mechanism carefully tuned for the target information.
- The main decision maker in the WAIS system is the user; only the user can determine the relevance between a piece of information and the query used to find it. WAIS clients usually present a ranked list of objects to the user based on the given search criteria.
- There are WAIS gateways for Gopher and WWW.

World Wide Web

SLAC

NIKHEF DESY FNAL YOU CERN

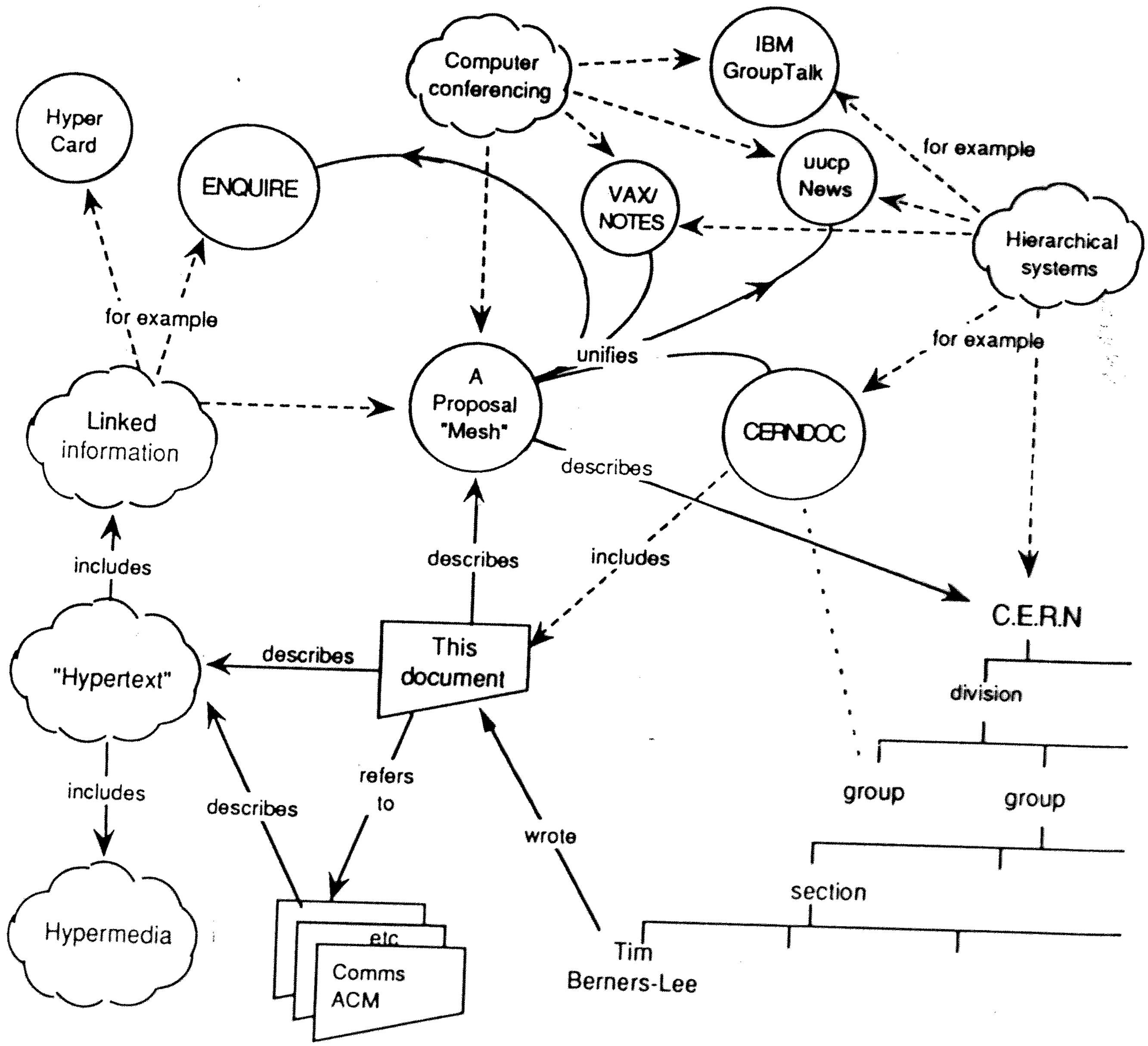
Global Hypertext

Information Management: A Proposal

Tim Berners-Lee, CERN

March 1989, May 1990

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system.



World Wide Web

CERN SLAC NIKHEF DESY FNAL YOU

Bebo White
SLAC, CERN
SHARE 80
March 4, 1993

SLAC World Wide Web Information

File Documents Navigate Customize Help

SLAC World Wide Web Information

The SLAC World Wide Web (WWW) service provides access to a wide range of information from a variety of sources both at SLAC and elsewhere. Information is provided in the form of hypertext which allows users to follow pointers from one item to other related items, and thus to access the information they need in a simple intuitive manner. Information is accessed using the TCP/IP network and can actually reside anywhere in the world, although the user does not need to know where the information comes from.

A number of different programs (called browsers) are available to access WWW information for different platforms including VM, VMS, Unix and NeXT.

SLAC information

SLAC SPIRES	Information is available on SLAC people (BINLIST), HEP people (HEPNAMES), HEP publications, as well as many additional topics.
Seminars	Seminars at SLAC and nearby, today, tomorrow, this week, next week, anytime.
SLD	Information specific to the SLD experiment.
FreeHep	Information on software available from HepLib.
VMS Help	Help information from SLACVX.

Other Information

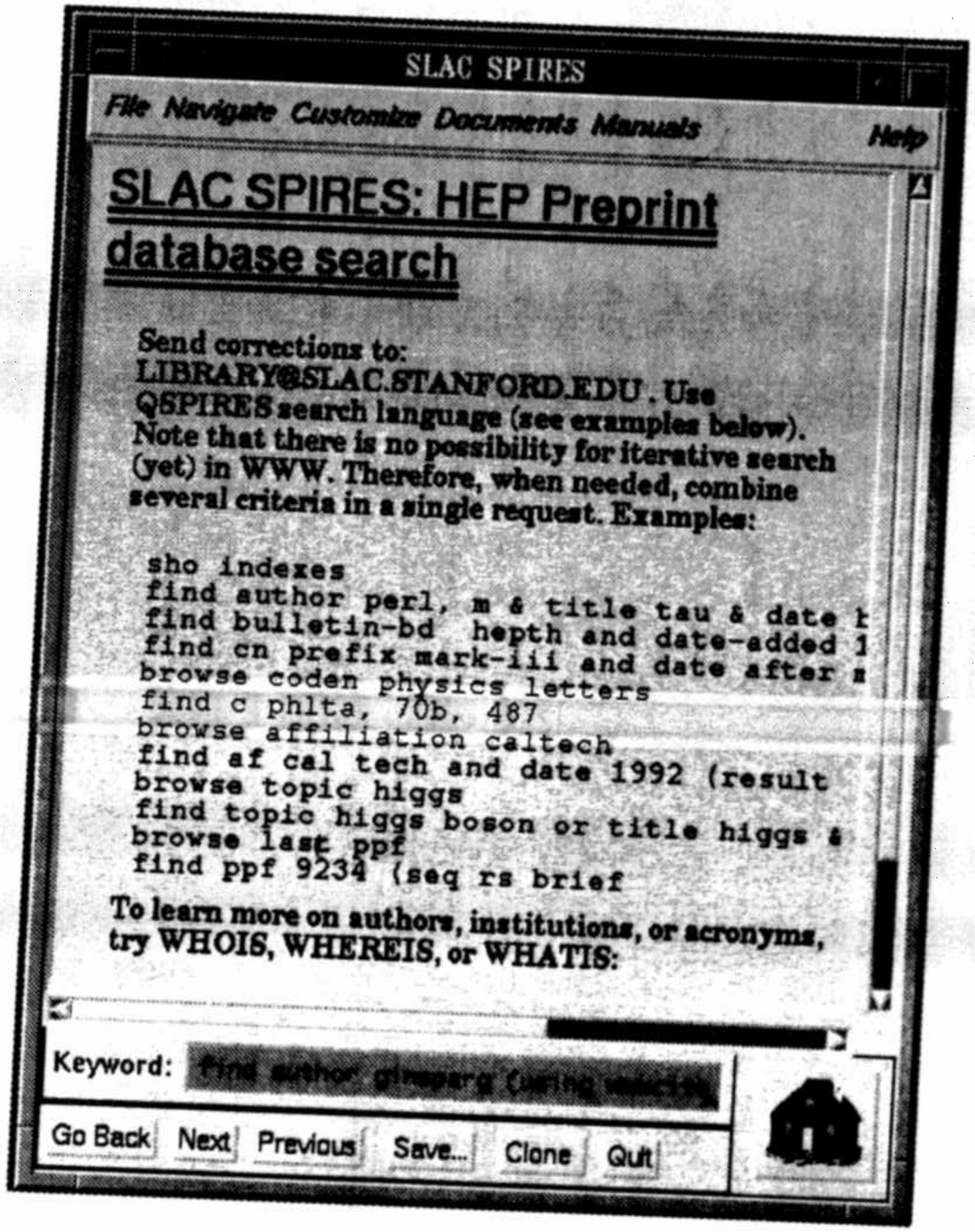
See also HEP information (CERN, DESY etc.), academic information, APS News.

Support

The WWW at SLAC is supported by the SLAC WWWizards, to whom questions – comments – complaints etc. should be addressed. The WWW project was initiated and is supported by a group based chiefly at CERN.

Keyword:

Go Back | Next | Previous | Save... | Clone | Quit


SLAC SPIRES
File Navigate Customize Documents Manuals
Help
SLAC SPIRES: HEP Preprint database search
Send corrections to: LIBRARY@SLAC.STANFORD.EDU . Use QSPIRES search language (see examples below). Note that there is no possibility for iterative search (yet) in WWW. Therefore, when needed, combine several criteria in a single request. Examples:
sho indexes
find author perl, m & title tau & date
find bulletin-bd hepth and date-added
find cn prefix mark-iii and date after
browse coden physics letters
find c phlta, 70b, 487
browse affiliation caltech
find af cal tech and date 1992 (result
browse topic higgs
find topic higgs boson or title higgs &
browse last ppf
find ppf 9234 (seq rs brief
To learn more on authors, institutions, or acronyms, try WHOIS, WHEREIS, or WHATIS:
Keyword:
Go Back
Next
Previous
Save...
Clone
Quit

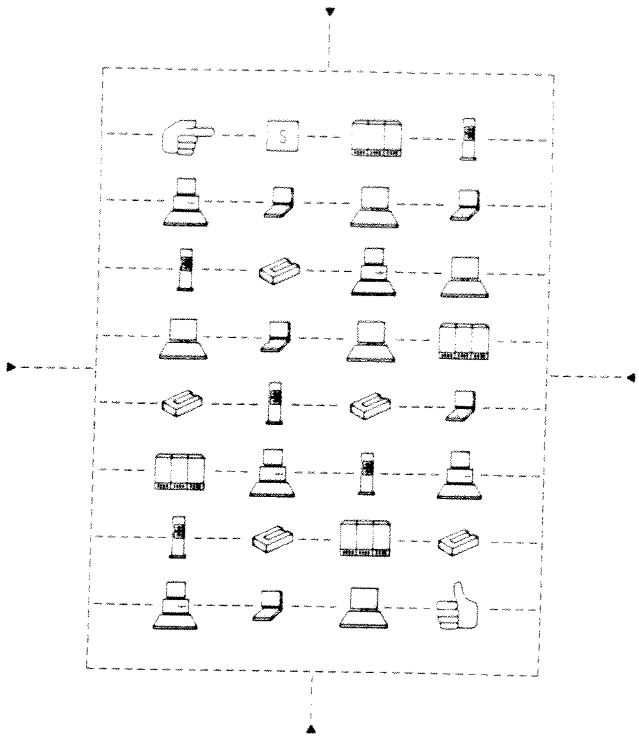
World Wide Web

SLAC NIKHEF DESY FNAL YOU CERN

Global Hypertext

World Wide Web

Bebo White
SLAC, CERN



SHARE

Academic & Research Technologies Project Newsletter

SHARE Winter 1993 Meeting

San Francisco, California

February 28—March 5, 1993

In this Issue: *Page*

New Name for the University Project!

Ever get the feeling these days that changes will never stop coming? We do too, but some are actually for the better. So the time-honored name "University Project" passes into oblivion in favor of the new techno-correct name *Academic & Research Technologies Project* — *ART* for short. In many ways, the new name better reflects who we are and what we do at SHARE. Since we've long had a fondness for sessions presenting the latest application of computing tools to research and education, this change well suits the project's interests. Look for us now in the Management Division under *A* rather than *U* (jumping from the end of the alphabet to the beginning must have some advantage!). And consider, now, how much better this will look on your resumé when you volunteer for Project work!

Highlights of the Academic and Research Technologies Project Sessions

The Academic and Research Technologies Project is sponsoring a dozen sessions at SHARE 80 in San Francisco that explore two main themes— the leading edge in ***networked academic information systems and services,*** and ***future directions in campus computing environments*** for the remainder of this decade and the century to come. Topping it all off is the pièce de résistance —a presentation on computing and social responsibility by a representative of the national nonprofit public interest organization, Computer Professionals for Social Responsibility. This article will highlight these and other sessions that you just won't get from any other SHARE project.

**All ART sessions will be in the Parc 55 Hotel,
55 Cyril Magnin Street,
across the street from the Hilton.**

Social Responsibility at SHARE!

My historical memory may be limited, but, since the early 1980s when this project manager appeared on the scene at SHARE, there has not been a session expressly devoted to the *political issues surrounding technology policy, the accelerating information revolution, and where computing professionals fit in.* from a progressive point of view How can we work to dispel popular myths about the infallibility of technological systems? How do we translate our vision of technology's promise into a truly democratic reality?

Evelyn Pine, Managing Director of **Computer Professionals for Social Responsibility**, a national nonprofit alliance of computer scientists and information industry specialists, will discuss the impact that we can have collectively and individually in shaping the technology policy of the future. (**M556**: Wednesday, 3:00 p.m., Parc Ballroom II).

The Future of Campus Computing

Two sessions will provide perspective on campus computing environments of the future.

Sandy Moy, SHARE's immediate past President, will discuss the changing paradigm that characterizes academic computing, including changes in economics, technology, topology, and useability of computing resources, accompanied by a dramatic increase in computer literacy. It is our challenge to try to understand the effect that these have on priorities in traditional computing services organizations. Join us to hear her unique perspective and to share your ideas! (**M552:** Tuesday, 9:30 a.m., Barcelona I)

Jack McCredie, a past president of EDUCOM and a pioneer in campus computing planning, will address the shifting paradigm from the perspective of the dramatic changes in computing resources devoted to instruction, research, and administration. How do we anticipate and plan for these rapid changes? (**M555**: Wednesday, 11:00am, Parc Ballroom II).

Networking Sessions at SHARE 80

By popular request, **Larry Snodgrass**, of the BITNET BITNIC, begins our SHARE week by describing how to use the BITNET II protocol and an Internet link to replace a traditional bisync BITNET link, reduce costs (by eliminating one connection), and increase throughput (by using the higher-speed Internet connection for both networks). Those facing budget cutbacks will find this an important session to attend. (**M551**: Monday, 4:30 p.m., Michelangelo).

We Need Your Help

Please contact one the Project Officers if you can help with session planning, chairing sessions, distributing and collecting session evaluation cards, or other project work. We have had a lot of turnover in producing the newsletter and would love to have a stable home for it. This newsletter was produced using MS Word for Windows. It was printed from a Mac on an Apple Laserwriter NT (because the PS/2 couldn't handle the graphics!). Virtually any system will do. Can you help with this?

Mastering the Internet

Success in presenting our first *Mastering the Internet* series at SHARE 79 spurred us to build an exciting new slate of sessions occupying much of Thursday of SHARE week. We had packed audiences for all three sessions and favorable reviews. Special thanks go to Richard Henderson of the UC Berkeley Haas School of Business for organizing the first set and several of this set. This round, we've expanded to four sessions, changing and adding new topics.

Kicking off the series is *Network Architecture and Basic Services,* a presentation by Georgia Tech's network wizard **David Buechner** that focuses on making effective use of both BITNET and the Internet via a comprehensive overview of network services. (**M558**: Thursday, 11:00 a.m., Parc Ballroom I).

Armed with a working knowledge of the underlying architecture, you'll be treated to a survey by UCB's **Bernard Aboba** of the literally hundreds of library catalogs, databases, and specialty resources currently available on the Internet, largely for free. (**M559**: 11:00 a.m., Parc Ballroom I).

Two new sessions have been added just for the SHARE 80 series. **Tim Howes** of ITD Systems/University of Michigan, will describe efforts underway to create a world of interoperability of which we've only previously been able to dream. Imagine being able to query a database half way around the world to get the name and e-mail address of a colleague in another country. The X.500 protocol standard describes a mechanism for building a single, logically connected database of user and institution information whose distributed pieces exist on hundreds of computers interconnected by the Internet. Among other, the University of Michigan is taking the first steps as part of the PSI White Pages Pilot Project. (**M560**: Thursday, 1:30 p.m., Parc Ballroom I).

Bebo White of Stanford University winds up our Internet series by describing another Internet phenomenon, the World Wide Web. The WWW uses a hypertext user interface to tie together many varying information systems into a homogeneous, browsable, searchable "web." Originating in the high-energy physics community, the WWW will extend to multi-format delivery and collaborative authoring in the future. Don't miss this opportunity to understand and bring home information on fantastic networking advances. (**M561**: Thursday, 3:00 pm, Parc Ballroom I).

Please set Thursday aside for this information-packed series!

Infocal: Berkeley's Z39.50-based Campus Information Service

Several years ago, we heard about UC Berkeley's plan to build a campus-wide information server based on the emerging Z39.50 computer-to-computer protocol standard. **Margaret Baker**, manager of this project since its inception, will discuss today's *Infocal*, its interoperability with the library, full-text databases, and other types of campus online resources, and why this was necessary. (**M553**: Tuesday, 3:00 p.m., Parc Ballroom II).

Training for DP Professionals

Just *what is the appropriate* (university and/or other) *training* for a data processing professional? A University instructor, **Dr. Robert Rannie** (Northern Illinois University) and a corporate instructor, **Dr. James Leon** (Amdahl), will offer some answers from their particular perspectives. What are your sentiments? (**M562**: Thursday, 8:00 am, Parc Ballroom I).

Issues and Problems / Requirements

We'll have our standard *Issues and Problems* session (**M556**: Wednesday, 4:30 p.m. Sienna) and *Requirements* (**M554**: Tuesday, 4:30 p.m., Michelangelo). The following article provides more detail on **Requirements.**

Mary Engle
Academic and Research Technologies
Project Manager

Requirements and the HESC

John Nolan, Cleveland State University (CSU), was named the *Deputy Project Manager for Requirements.* At the last session we noted several old requirements with low actvity. We would like your opinion on whether these are obsolete or if they need to be updated and resubmitted. John will report on these at the requirements session.

We will discuss the **HESC** as part of the Project Requirements session (**M554:** Tuesday , 4:30 p.m. Michelangelo Room, 4th Floor of the PARC 55). **Fred Dwyer**, IBM's Administrator for the HESC program, will participate in this discussion. This is our chance to provide valuable feedback on the HESC to those involved in the decision-making process at IBM. Come prepared to let IBM know what we really need from IBM and the HESC program to maximize its benefits to both the University community and International Business Machines.

Please come to share your opinions!

We would also like to carry on requirements discussion and voting over BITNET, but there are confidentality concerns in handling IBM Requirements over the net. John will look into this and report to the Project as soon as possible.

The IBM Requirements process is undergoing an update. From the Requirements process, IBM needs to understand *what the problem is, why is it a problem, how it effects the customer,* and *how the customer benefits.* Some categorys of benefits include: Business Solutions, Investment Protection, End User Productivity, Growth Enablement, and Systems Management. You may have noticed this new terminology in recent major announcement letters.

John Nolan
Deputy Project Manager—Requirements

Changing of the Guard

SHARE 80 will be **Mary Engle**'s last SHARE conference. Mary has been involved with the Project leadership for five years, and has been its PM for most of that time. Thanks for a job well done.

Diana D'Angelo leaves the post of Deputy Project Manger. Diana served a term of nearly five years as Deputy Project Manager, and created and maintains the current Project mailing list. Thanks to Diana! We will certainly miss her energy and wonderful spirit.

Stan Yagi, of Queen's University in Kingston, Ontario, will become Project Manager beginning with SHARE 81. Stan comes from the Integrated Technologies Division, where he was Division Manager. We wish him the best of luck! Please continue to offer your time and energy to help Stan keep this project vital, productive, and attractive to newcomers.

Project Leadership Opportunities

The ART Project will be recruiting the following positions:

- Deputy Project Manager to replace Diana D'Angelo.
- Newsletter Editor

The Project also needs a new home for its mailing list database. The database is currently in a Nomad database on an IBM 3090 at Michigan State, and can be output in ASCII if needed. We update it after each meeting, and twice a year print labels and the mailing lists that we circulate at our sessions. If you can help by providing a new home for the mailing list and produce the printed products twice yearly, please contact one of the Project officers listed on the last page.

Management Trade Show

The Management Division will sponsor a "trade show" at SHARE 80 on Sunday night from 6:00 p.m.—8:00 p.m. at SCIDS. Its purpose is to acquaint newcomers and those unfamiliar with the Management Division with its projects and their activities. Each project will have a table with descriptive project literature, grids, and warm bodies to greet newcomers. If you can assist at the ART Project table, please join us at SCIDS on Sunday night anytime between 6 and 8 p.m. Or join us there to find out about the week's activities. The trade show will replace the regular ART Project opener normally held during this time.

ART Project Dinner

The Academic and Research Technologies Project will gather for a dutch-treat dinner after the end of the Management Trade Show at SCIDS on **Sunday night**. Even if you don't make it in time to hang around the table, please join us for dinner at **8:00 p.m**. **Meet at the ART table at SCIDS**. If you ***know*** you can make it, please let Mary Engle know asap by email (meeur@uccmvsa.bitnet) so she can make some reservations.

Restaurant Lists

The ART Project will have San Francisco restaurant lists available at the project table on Sunday night at SCIDS between 6 and 8 p.m. Please stop by to pick one up and greet your university colleagues.

Project Officers

Project Manager:
Mary Engle (UCA)
University of California
Division of Library Automation
300 Lakeside Drive, 8th floor
Oakland, CA. 94612-3550
(510) 987-0563
E-mail: meeur@uccmvsa.bitnet

Deputy Project Manager--Requirements
John Nolan (CSU)
Cleveland State University
1820 E. 22nd Street
Cleveland, OH 44144
(216) 687-21509
E-mail: s0001@csuohio.bitnet

Deputy Project Manager-Program Planning:
Richard Henderson (UBB)
University of California, Berkeley
Haas School of Business
350 Barrows Hall
Berkeley, CA. 94720
(510) 642-6913
E-mail: richard@haas.berkeley.edu

Deputy Project Manager:
Diana D'Angelo (MSU)
Michigan State University
315 A Computer Center
East Lansing, MI 48824-1042
(517) 353-1861
E-mail: dianad@msu.bitnet

Deputy Project Manager :
(PM beginning SHARE 81)
Stan Yagi (QK)
Queens University
Computing and Communications Services
Dupuis Hall
Kingston, Ontario K7L 3N6
CANADA
613) 545-2051
yagi@qucdn.queensu.ca

Project Officer:
John Sheshene (SJU)
St. John's University
Academic Computing Services
Perboyre Hall B14
Jamaica, NY 11439
(718) 990-1674
E-mail: sheshene@sjuvm.bitnet

IBM Representative:
Bob Jones (IBM)
IBM
Academic Information Systems (ACIS)
472 Wheelers Farms Road
Milford, CT 06460
(203) 783-7841

UNIVERSITY PROJECT SESSIONS

SHARE 80 — San Francisco, California

February 28 — March 5, 1993

Time	Monday 3/1	Tuesday 3/2	Wednesday 3/3	Thursday 3/4	Friday 3/5
8:00 – 9:00 a.m.				**M562:** Appropriate Training for a DP Professional (Parc 55, Parc Ballroom I)	8:30 a.m. **University Project Breakfast Wrapup** (Laurie's Diner, 336 Mason)
9:30 – 10:30 a.m.	General Session:	**M552: Campus Computing Environments of the 1990's** (Parc 55, Barcelona I)		***Mastering the Internet*** **M558: Network Architecture and Basic Services** (Parc 55, Parc Ballroom I)	
11:00 – 12:00 p.m.	Management Division Opening Session		**M555: Planning Future Directions for Computing on Campus** (Parc 55, Parc Ballroom II)	**M559: Network Information Resources** (Parc 55, Parc Ballroom I)	
1:30 – 2:30 p.m.				**M560: Using the OSI X.500 Directory Services: the White Pages Project** (Parc 55, Parc Ballroom I)	
3:00 – 4:00 p.m.		**M553: Infocal: UC Berkeley's Z39.50-based Campus Information Service** (Parc 55, Parc Ballroom II)	**M556: Computing Technology and Democracy** (Parc 55, Sienna)	**M561: World Wide Web: The Universe of Information** (Parc 55, Parc Ballroom I)	
4:30 – 5:30 p.m.	**M551: BITNET II: Combining Your BITNET and Internet Links** (Parc 55, Michelangelo)	**M554: Academic and Research Technologies Project Requirements (General and HESC)** (Parc 55, Michelangelo)	**M557: Academic Computing Issues and Problems Free-for-All** (Parc 55, Sienna)		
6:00 p.m.	***Sunday: SCIDS Table***				

S aker Name: Bebo White

Session Number: M561

Title: Mastering the Internet Part 4:

Abstract:

The W3 initiative ties together many varying information systems into a homogeneous browsable, searchable "web." By combining a hypertext user interface with index queries, the data model allows almost any existing information system to be represented in the web. The simple intuitive interface gives the impression of a single source when data is in fact furnished by FTP, Gopher, WAIS and local W3 servers. A practical collaboration rather than research project, W3 has seen success as a distributed information system, and will in the future extend to collaborative authoring and multi-format delivery.

This talk will outline the techniques used, describe methods of making data available on the "web", describe the various browsers available for different platforms, discuss the extent of the web in various research communities, and share future plans.

Day/Time/Location: THURSDAY, 3:00 PM
PARC FIFTY FIVE, FOURTH FLOOR, PARC BALLROOM I

AV Equipment: Flip Chart, Lavalier Microphone, Overhead Projector

Capacity/Setup: 100, Theater

Session Chairman: Linda Littleton
Penn State University
Computation Center
University Park PA 16802
(814) 865-0819

Project Manager: Mary Engle (510) 987-0563

Tony Johnson

Tel: (415) 926 2278, TONYJ@scs.slac.stanford.edu.

Designer of MidasWWW . Boston University, collaborating with SLAC, SSC, etc. A SLAC server expert and a WWWizard .

Marc Andreesen

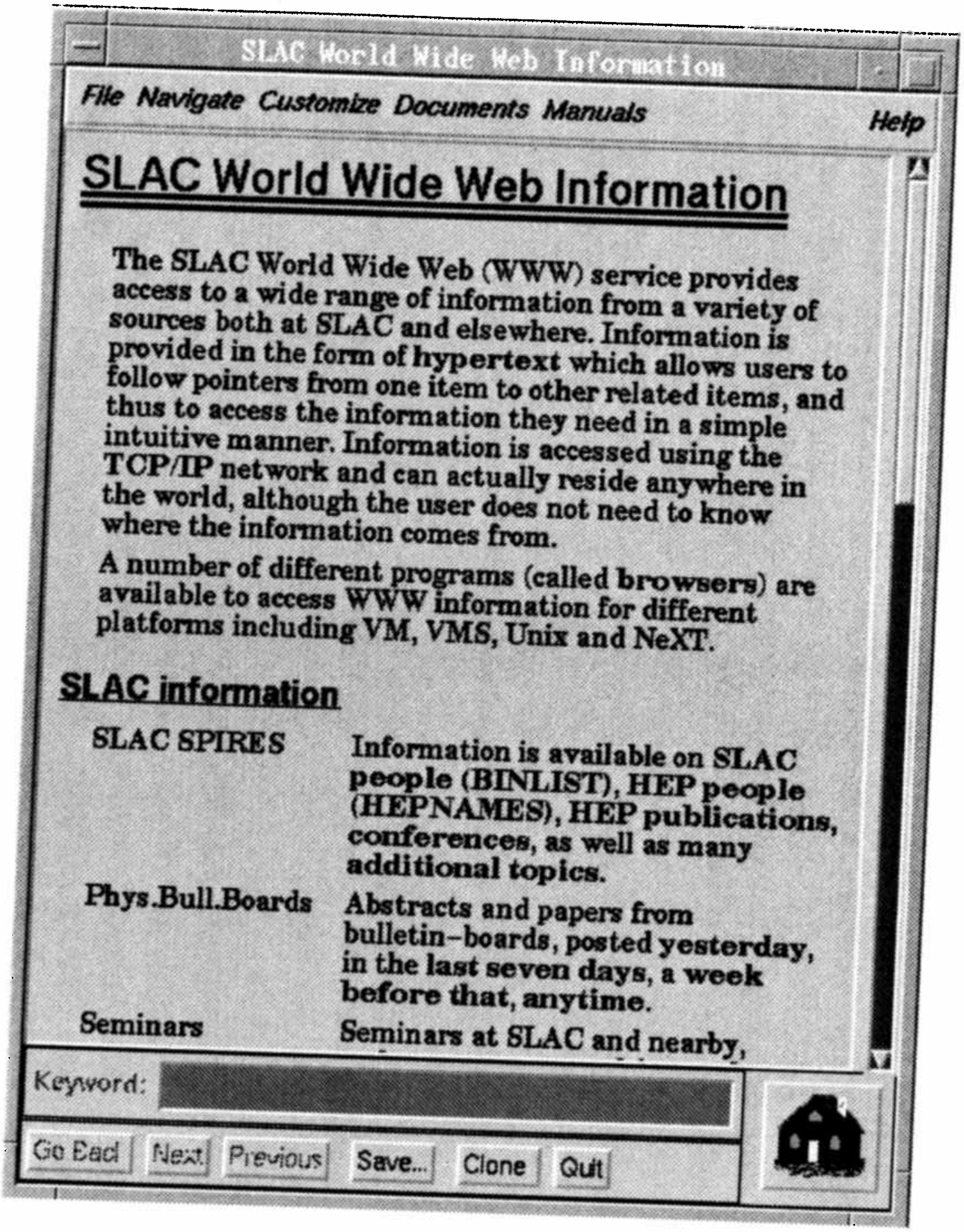
National Center for Supercomputing Applications (NCSA), Urbana Champagne, IL, USA. Design lead and co-developper of XMosaic . <marca@ncsa.uiuc.edu>. (more

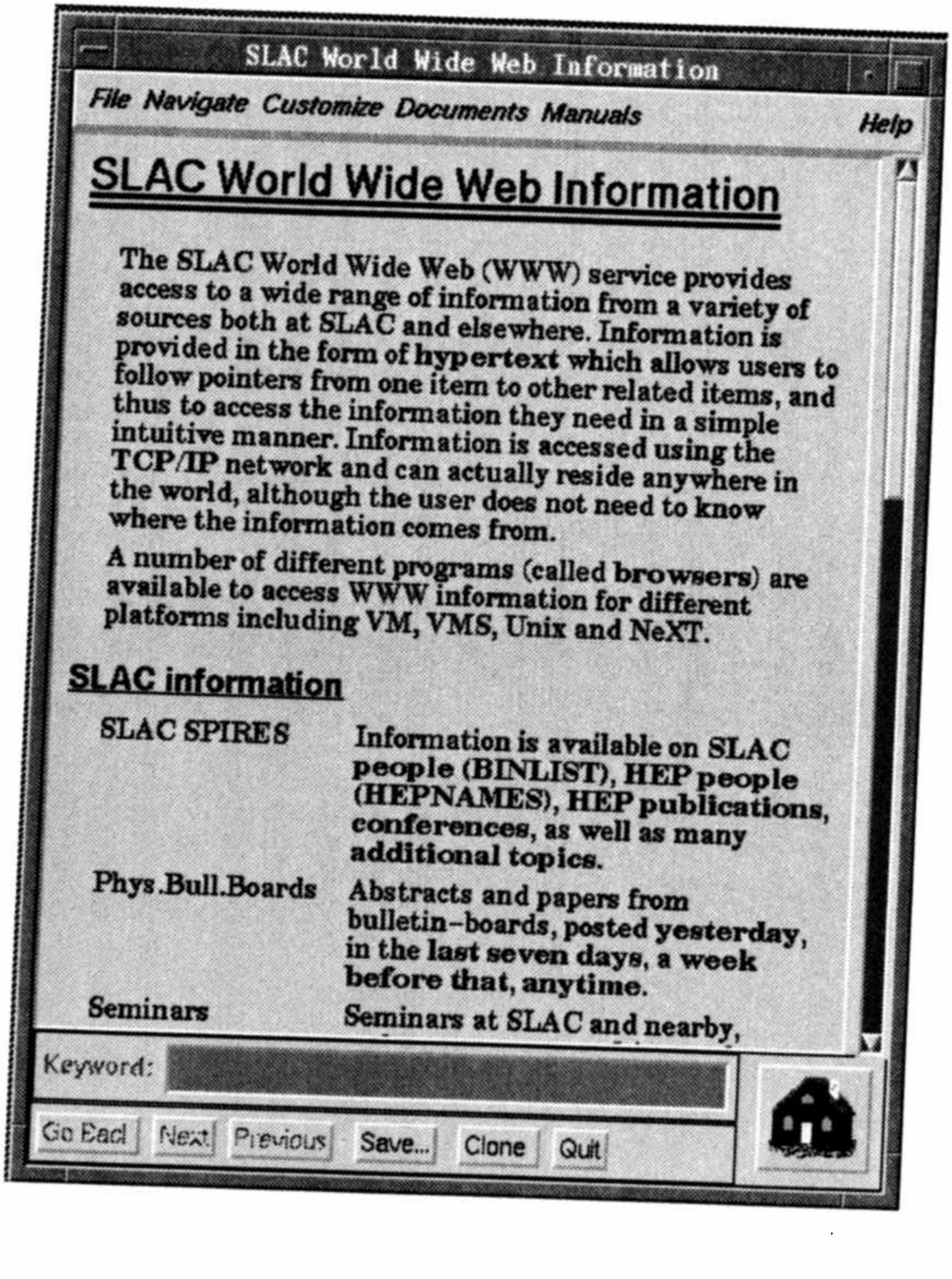
Pei Wei

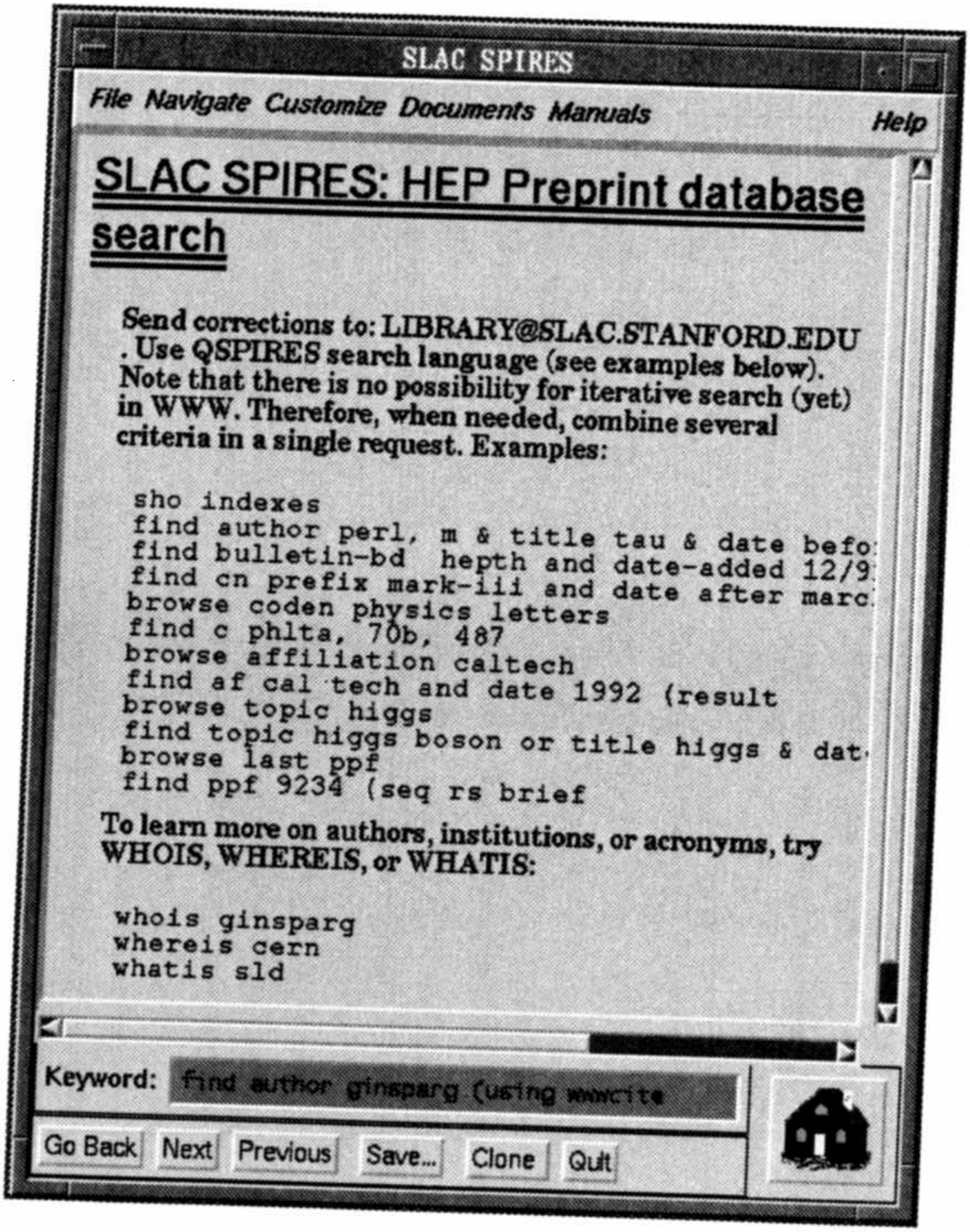
Pei is the author of " Viola", a hypertext browser, and the ViolaWWW variant which is a WWW browser. He was at the University of California at Berkeley, Experimental Computing Facility, now full time with O'Reilly and Associates, Sebastopol, CA, USA. Email: wei@xcf.berkeley.edu

Thomas R Bruce

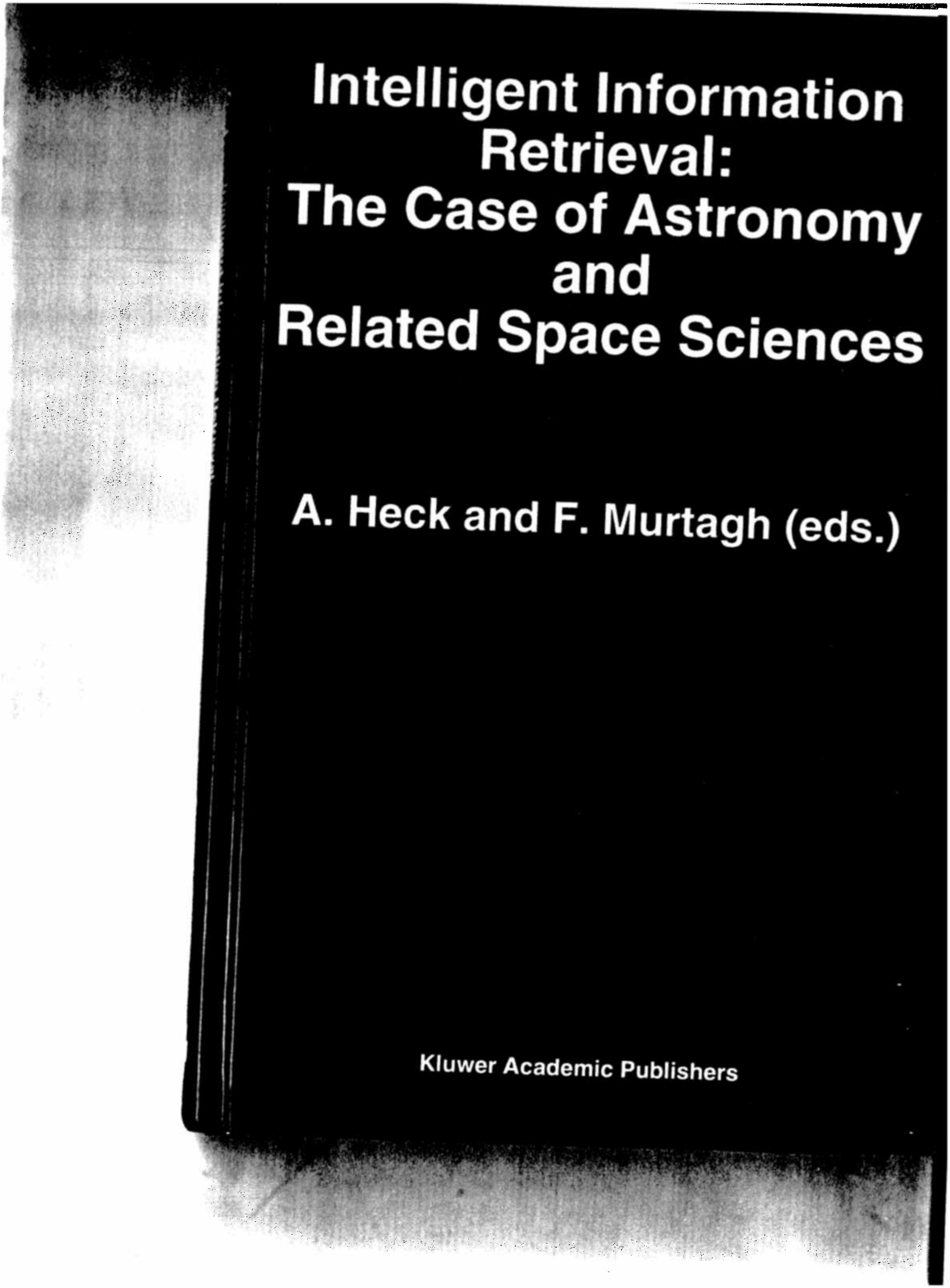
Formerly a staff member in charge of computer operations at the Cornell Law School, Tom is now a research associate working on a variety of projects invol the dissemination of legal information on the Internet. He is the author Cello all-singing, all-dancing WWW browser for Microsoft Windows. E-mail:tom@law.mail.cornell.edu.


SLAC World Wide Web Information
File Navigate Customize Documents Manuals
Help
SLAC World Wide Web Information
The SLAC World Wide Web (WWW) service provides access to a wide range of information from a variety of sources both at SLAC and elsewhere. Information is provided in the form of hypertext which allows users to follow pointers from one item to other related items, and thus to access the information they need in a simple intuitive manner. Information is accessed using the TCP/IP network and can actually reside anywhere in the world, although the user does not need to know where the information comes from.
A number of different programs (called browsers) are available to access WWW information for different platforms including VM, VMS, Unix and NeXT.
SLAC information
SLAC SPIRES
Information is available on SLAC people (BINLIST), HEP people (HEPNAMES), HEP publications, conferences, as well as many additional topics.
Phys.Bull.Boards
Abstracts and papers from bulletin-boards, posted yesterday, in the last seven days, a week before that, anytime.
Seminars
Seminars at SLAC and nearby,
Keyword:
Go Back
Next
Previous
Save...
Clone
Quit


SLAC World Wide Web Information
File Navigate Customize Documents Manuals
Help
SLAC World Wide Web Information
The SLAC World Wide Web (WWW) service provides access to a wide range of information from a variety of sources both at SLAC and elsewhere. Information is provided in the form of hypertext which allows users to follow pointers from one item to other related items, and thus to access the information they need in a simple intuitive manner. Information is accessed using the TCP/IP network and can actually reside anywhere in the world, although the user does not need to know where the information comes from.
A number of different programs (called browsers) are available to access WWW information for different platforms including VM, VMS, Unix and NeXT.
SLAC information
SLAC SPIRES
Information is available on SLAC people (BINLIST), HEP people (HEPNAMES), HEP publications, conferences, as well as many additional topics.
Phys.Bull.Boards
Abstracts and papers from bulletin-boards, posted yesterday, in the last seven days, a week before that, anytime.
Seminars
Seminars at SLAC and nearby,
Keyword:
Go Back
Next
Previous
Save...
Clone
Quit


SLAC SPIRES
File Navigate Customize Documents Manuals Help
SLAC SPIRES: HEP Preprint database search
Send corrections to: LIBRARY@SLAC.STANFORD.EDU . Use QSPIRES search language (see examples below). Note that there is no possibility for iterative search (yet) in WWW. Therefore, when needed, combine several criteria in a single request. Examples:
sho indexes
find author perl, m & title tau & date befo
find bulletin-bd hepth and date-added 12/9
find cn prefix mark-iii and date after marc
browse coden physics letters
find c phlta, 70b, 487
browse affiliation caltech
find af cal tech and date 1992 (result
browse topic higgs
find topic higgs boson or title higgs & dat
browse last ppf
find ppf 9234 (seq rs brief
To learn more on authors, institutions, or acronyms, try WHOIS, WHEREIS, or WHATIS:
whois ginsparg
whereis cern
whatis sld
Keyword: find author ginsparg (using wwwcite
Go Back
Next
Previous
Save...
Clone
Quit

©1993


Intelligent Information Retrieval:
The Case of Astronomy and Related Space Sciences
A. Heck and F. Murtagh (eds.)
Kluwer Academic Publishers

INTELLIGENT INFORMATION RETRIEVAL: THE CASE OF ASTRONOMY AND RELATED SPACE SCIENCES

Edited by

A. HECK

Observatoire Astronomique, Strasbourg, France

and

F. MURTAGH

ESO ST-ECF, Garching bei München, Germany

KLUWER ACADEMIC PUBLISHERS

DORDRECHT / BOSTON / LONDON

Library of Congress Cataloging-in-Publication Data

Intelligent information retrieval : the case of astronomy and related
space sciences / edited by A. Heck and F. Murtagh.
p. cm. -- (Astrophysics and space science library ; v. 182)
Includes index.
ISBN 0-7923-2295-9 (alk. paper)
1. Astronomy--Data bases. 2. Space sciences--Data bases.
3. Information storage and retrieval systems--Astronomy.
4. Information storage and retrieval systems--Space sciences.
5. Expert systems (Computer science) I. Heck, A. (André)
II. Murtagh, Fionn. III. Series.
QB51.3.E43I56 1993
025.06'5005--dc20

93-19301

ISBN 0-7923-2295-9

Published by Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Kluwer Academic Publishers incorporates
the publishing programmes of
D. Reidel, Martinus Nijhoff, Dr W. Junk and MTP Press.

Sold and distributed in the U.S.A. and Canada
by Kluwer Academic Publishers,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers Group,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands.

Printed on acid-free paper

Printed in the Netherlands

Table of Contents

Chapter 10

WorldWideWeb (WWW)

Bebo White

Stanford Linear Accelerator Center
P.O. Box 4349, MailStop 97
Stanford, CA 94309 (USA)
Email: bebo@slac.stanford.edu

10.1 Introduction

The term *hypertext* was coined by T.H. Nelson circa 1965 to describe text (later described as being words, pictures and sound) which is not constrained to be linear (i.e. sequential; Kahn et al, 1988; Nelson, 1990). In particular, hypertext can be viewed as text which contains links to other texts. A *hypertext document* can therefore be thought of as a collection of texts and the links between them which can logically be viewed as a single entity. There is no requirement that the constituent parts of a hypertext document must be stored as a single entity. An *anchor* within a hypertext document is defined as some collection of text which is the source or destination of a link. Figure 10.1 provides a simplistic view of the construction of a hypertext document.

In 1965 hypertext was an interesting philosophical concept. Meaningful applications of these concepts were difficult to conceive. However, in the 1990s high-speed networks and information servers have made hypertext an important concept in the design of information retrieval systems. The definition of network protocols specifically designed to emulate hypertext links have made hypertext documents a reality.

10.2 What is WorldWideWeb?

WorldWideWeb (WWW or W3) has introduced to information retrieval on the Internet, the concept of *hypertext*. Rather than rely upon an existent hierarchy of information or a keyword-based search, this method attempts to link information in a manner which

A. Heck and F. Murtagh (eds.), Intelligent Information Retrieval: The Case of Astronomy and Related Space Sciences, 127–133.

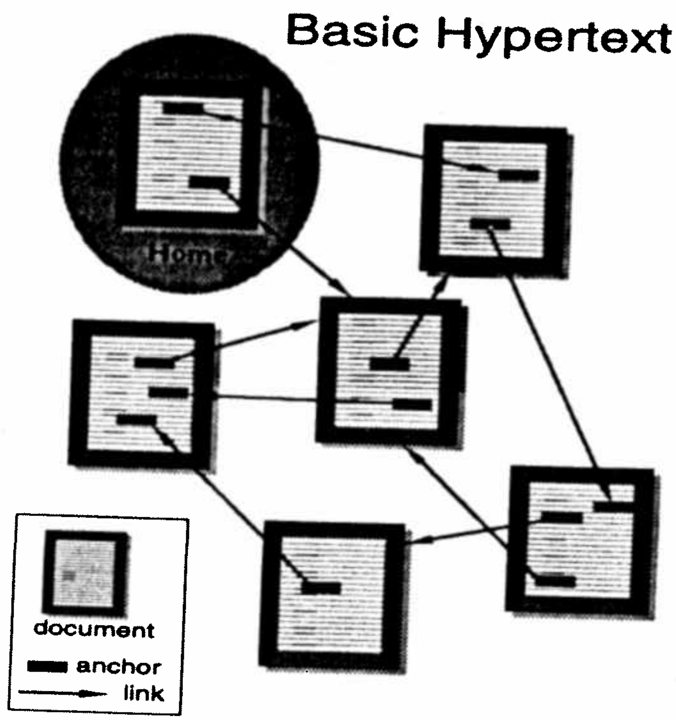


Figure 10.1: An illustration of hypertext.

mimics the human association of ideas. As a result, the associative links, access paths, etc. between information and/or information sources more closely resemble a spider's *web* rather than the conventional tree-like or directory-like structure. The topology of the Internet nodes within this information network can also be thought of as a similar web.

The WWW user is therefore able to follow a multiplicity of paths (even circular) in order to find the information required. In the meantime, such paths may allow the user access to additional similar or dissimilar information, associated information (e.g. definitions, etc.), and access to other information retrieval systems.

10.3 WWW Features

10.3.1 How does WWW look to the network?

Like a number of other information retrieval systems (e.g. Gopher and WAIS), WWW presents network-distributed information. The WWW architecture involves two programs, a client "browser/reader" and a server, communicating across the network. The "information bus" which connects clients consists of a uniquely defined set of standards

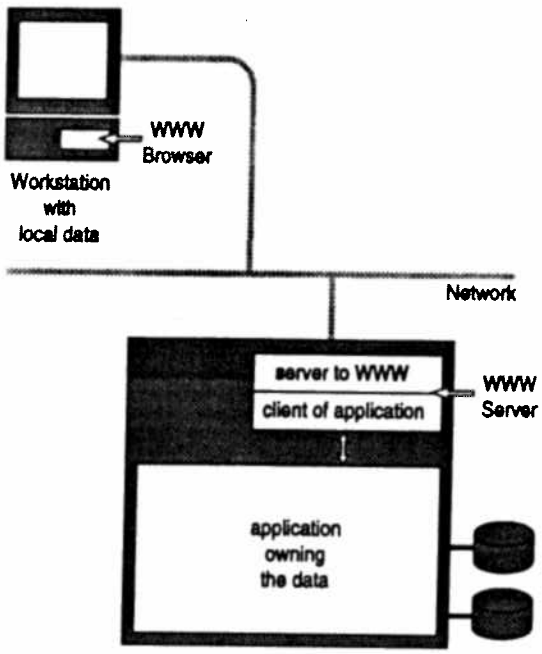


Figure 10.2: The browser and server relationship.

and conventions. The three conventions of this "information bus" are Universal Resource Locators, a set of protocols, and a set of data formats. At this level, the user-perceived "web" is composed of named documents some of which may in actuality be searchable indices. The browser can present the documents and the relationships between them in a manner suited to the needs and/or desires of the user. The server is free to display the documents either by sending real files (either with or without hypertext links) or by generating virtual hypertext dynamically in response to a request to some service available to the server. This is how a *gateway* server can provide a hypertext image of all the data in another system. Therefore, information providers, such as some of the others described in chapters 7, 8 and 9, are not forced to change their interfaces or convert information. This process is handled by the gateway server. Figure 10.2 provides a simplistic view of this browser/network/server relationship.

10.3.2 How does WWW look to the user?

To a user, WWW consists of two fundamental operations – following a hypertext link and performing a keyword indexed database search. The usual WWW client interface (i.e. browser) resembles a menu (commonly called a *page*) which contains active buttons (the hypertext links) or keys. The entry-level page at a particular node is commonly referred to as that node's *front page*. On platforms which allow hypertext-processing (e.g.

mouse-driven systems), the user "points and clicks" on these buttons in order to follo a link.

The default "line-mode" or TTY browser presents these links as numeric menu item The number of the link to be followed is typed by the user on the command line. Consi tent with WWW's hypertext metaphor, none of these buttons/links need be sequenti or hierarchical in their relationship to one another.

A link can be a dynamic request to an index server. Keywords for index searches ar usually typed in a window especially provided for that purpose. The Midas interfac is one such which may be used for keyword searches, with the index server being th Gopher Server.

Consistent with WWW's distributed information model, the links followed fron a page may be on many different machines anywhere on the network. As the use pushes buttons and browses one file after another, he/she may be actually queryin information servers throughout the network, exploring therefore a large informatior *web*. The traversal of this web is usually transparent to the user in that WWW handles al of the necessary network protocols.

10.3.3 General features of WWW

The general features of WWW may be summarized as follows (Berners-Lee et al., 1992):

- WWW Can Insure Access to Most Current Information

  Information need only to be represented once, as a reference to the original/master source document may be made instead of making a local copy.

  Links allow the topology of the information to evolve, so modeling the current state of a topic area of interest is possible at any time without constraint.

- WWW Optimizes the Use of an Information Source

  The user-available information web stretches seamlessly from small personal notes on local workstations to large databases anywhere on a network.

  Indices are documents, and so may themselves be found by searches and/or following links. An index is represented to the user by a *cover page* that describes the data indexed and the properties of the search engine.

  The documents in the web do not have to exist as files; they can be *virtual* documents generated by a server in response to a query or document name. They can therefore represent views of databases, or snapshots of changing data (e.g. time-dependent data, jointly-authored works-in-progress).

10.4 WWW Conventions

As described earlier, the success of WWW operation depends upon three conventions – a set of protocols, Universal Resource Locators and a set of data formats.

10.4.1 Protocols

A number of existing protocols, and one new protocol, form the set with which a WWW browser is equipped. Standard protocols used are:

- FTP

  FTP (File Transfer Protocol) allows access to libraries of documents, software, etc. available on ftp servers which allow "anonymous" access.

- NNTP

  NNTP (Network News Transfer Protocol) allows access to Usenet Netnews servers and the display of news groups and articles.

A WWW browser is also equipped to handle two of the protocols of other information retrieval systems discussed in this book:

- WAIS

  Allows gateway access to WAIS (Wide Area Information Servers). WAIS is described in detail in chapter 8.

- Gopher

  Allows gateway access to the Internet Gopher system. Gopher is described in more detail in chapter 9.

A new protocol, HTTP (Hypertext Transport Protocol), was developed to allow document retrieval and index search (Berners-Lee, 1991). This protocol coordinates communication between WWW browsers and servers and is easily implemented in shell script, perl, C or other programming languages. HTTP is designed to be as efficient and compact as possible so as to minimize round trip delays between the browser and server nodes. A great deal of its flexibility lies in the Universal Resource Locator.

10.4.2 Universal Resource Locator

The Universal Resource Locator (URL) is used to define the essential information in a link. In earlier WWW documentation, the URL was referred to as the UDI (Universal Document Identifier), but has been changed to reflect the widening variety of information resources now available. The URL is designed to be compact and printable (i.e. contain no special characters). The URL contains fields identifying the applicable protocol, a server specification (usually just the fully-qualified Internet name), a port address, a document to be identified on that server, and a search to be performed if necessary. It may optionally contain a field which specifies a particular part of a document to be selected (an *anchor*) when the document is presented.

10.4.3 Data formats

For WWW the lowest common denominator for data is plain text. This format allows the ready display of simple documents and text output of index searches.

WWW servers may also generate simple hypertext documents written in a specially-designed text-processing language, *HTML*. HTML (Hypertext Markup Language) is a subset of the popular SGML (Standardized General Markup Language). It has a few simple formatting options (tags) which allow it to be effectively used for on-line documentation as well as menus and search results. HTML permits browser-dependent formatting of hypertext documents and the definition of links. For example, in an HTML document a markup tag such as #anc would contain a URL (Universal Resource Locator) to create a hypertext link.

As formats for representing data continually evolve, access to files of these data types must become generally available. The WWW architecture proposes a negotiation between client and server to agree on a document format for transmission. *Format negotiation* is available for some formats in specific browsers. Formats of particular interest are graphic files (e.g. GIF, TIFF), text-formatted files (e.g. DVI, RTF) and sound files.

10.5 How to Get More Information on WWW

10.5.1 Public access to WWW

A fully implemented version of WWW is available via *telnet* to *info.cern.ch*. No username or password is necessary in order to use this service. This example, since it is via telnet, uses the linemode browser. The user will be able to explore the web, but will not be able to use a hypertext-based interface.

The *front page* of *info.cern.ch* is highly tailored for use by CERN, the European Particle Physics Laboratory in Geneva, Switzerland (though it changes quite often). However, links do provide access to a wide variety of other information sources, including WAIS and Gopher (discussed in chapters 8 and 9). For example, following links entitled *academic information* lead to resources in the areas of astronomy and astrophysics.

10.5.2 Obtaining WWW software

Access to WWW server and client software for a wide variety of platforms is available from numerous sources via anonymous ftp. Using archie (described in chapter 7), a user can identify a convenient ftp site from which to get the software by issuing the command `prog www` or `prog WWW`. Pointers to sources for all of the popular browsers (e.g. Viola, Midas, tkWWW, XMosaic, etc.) are also available. Archie can also be used to point the user to the wealth of WWW documentation that is also available.

The guaranteed latest WWW server software (for a wide variety of platforms) and some browser software (including the linemode browser) is available via anonymous ftp from `info.cern.ch`

10.5.3 Additional information sources

There are presently two electronic mailing lists available to those interested in WWW –

- www-announce

  This list keeps subscribers informed as to WWW progress, new software releases and new data sources.

- www-talk

  This list provides a mechanism for technical discussion for those users interested in developing WWW software, WWW protocol evolution and WWW-related technologies.

In order to subscribe to either of these lists, send electronic mail to `listserv@info.cern.ch`. The body of the mail should contain only the line:

```
add www-announce
```

or

```
add www-talk
```

The name and address occurring on the From: line of this mail will be added to the mailing list(s).

References

1. Berners-Lee, T., "HTTP As Implemented in WWW", CERN, December 1991.
2. Berners-Lee, T., Cailliau, R., Groff, J.-F. and Pollermann, B., "World-Wide Web: the information universe", *Electronic Networking*, Meckler, Spring, 1992.
3. Kahn, P.D., Pau, Meyrowitz, "Guide, HyperCard, and Intermedia: a comparison of hypertext/hypermedia systems", IRIS Technical Report 88-7, Brown University, Providence, RI, 1988.
4. Nelson, T., *Literary Machines*, 90.1, Mindful Press, Sausalito, CA, 1990.

```
\documentstyle[,11pt]{article}
%%%% \documentstyle[palatino,11pt]{article}
\topmargin=0truecm                % from top of phys. page to page's head
\oddsidemargin=1.5truecm          % left margin on odd-numbered pages
\evensidemargin=0.5truecm         % left margin on even-numbered pages
\textheight=21truecm              % height of the text on the page
\textwidth=15truecm               % width of the text on the page
%\def\arraystretch{1.2}           % stretch factor for array and tabular env.
\def\arraystretch{1.0}            % stretch factor for array and tabular env.
\listparindent=0em                % indentation of listitems
\topsep=0.5ex                     % separation at the beginning of a list
\itemsep=\topsep                  % separation of items
\begin{document}
\input{psfig}

\begin{center}

{\huge {\bf WorldWideWeb (WWW)}}

\bigskip

{\large {\bf Bebo White}}

\bigskip

Stanford Linear Accelerator Center, P.O. Box 4349, MailStop 97,
Stanford, CA  94309


{\tt bebo@slac.stanford.edu}

\end{center}

\section{Introduction}

The term {\bf hypertext} was coined by T.H. Nelson circa 1965 to describe
text (later described as being words, pictures and sound) which is not
constrained to be linear (i.e., sequential). [Kahn et al88][Nelson90].
In particular, hypertext can be viewed as text which contains links to
other texts. A {\bf hypertext document} can therefore be thought of as a
collection of texts and the links between them which can logically be
viewed as a single entity. There is no requirement that the constituent
parts of a hypertext document must be stored as a single entity.
An {\bf anchor} within a hypertext document is defined as some collection
of text which is the source or destination of a link. Figure~\ref{Hyper}
provides a simplistic view of the construction of a hypertext document.

\begin{figure}[t]
\psfig{figure=hyper.ps}
\caption{xxxx}
\label{Hyper}
\end{figure}

In 1965 hypertext was an interesting philosophical concept. Meaningful
applications of these concepts were difficult to conceive. However, in
the 1990's high-speed networks and information servers have made
hypertext an important concept in the design of information retrieval
```

systems. The definition of network protocols specifically designed to emulate hypertext links have made hypertext documents a reality.

\section{What is WorldWideWeb?}

WorldWideWeb (WWW or W3) has introduced to information retrieval on the Internet, the concept of {\bf hypertext}. Rather than rely upon an existent hierarchy of information or a keyword-based search, this method attempts to link information in a manner which mimics the human association of ideas. As a result, the associative links, access paths, etc. between information and/or information sources more closely resemble a spider's {\bf web} rather than the conventional tree-like or directory-like structure. The topology of the Internet nodes within this information network can also be thought of as a similar web.

The WWW user is therefore able to follow a multiplicity of paths (even circular) in order to find the information required. In the meantime, such paths may allow the user access to additional similar or dissimilar information, associated information (e.g., definitions, etc.), and access to other information retrieval systems.

\section{WWW Features}

\subsection{How Does WWW Look to the Network?}

Like a number of other information retrieval systems, (e.g., Gopher and WAIS), WWW presents network-distributed information. The WWW architecture involves two programs, a client "browser/reader" and a server, communicating across the network. The "information bus" which connects clients consists of a uniquely defined set of standards and conventions. The three conventions of this "information bus" are Universal Resource Locators, a set of protocols, and a set of data formats. At this level, the user-perceived "web" is composed of named documents some of which may in actuality be searchable indices. The browser can present the documents and the relationships between them in a manner suited to the needs and/or desires of the user. The server is free to display the documents either by sending real files (either with or without hypertext links) or by generating virtual hypertext dynamically in response to a request to some service available to the server. This is how a {\bf gateway} server can provide a hypertext image of all the data in another system. Therefore, information providers, such as some of the others described in this book, are not forced to change their interfaces or convert information. This process is handled by the gateway server. Figure~\ref{cliserv} provides a simplistic view of this browser/network/server relationship.

```
\begin{figure}[t]
\psfig{figure=cliserv.ps}
\caption{xxxx}
\label{cliserv}
\end{figure}
```

\subsection{How Does WWW Look to the User?}

To a user, WWW consists of two fundamental operations - following a hypertext link and performing a keyword indexed database search. The usual WWW client interface (i.e., browser) resembles a menu (commonly called a {\bf page}) which contains active buttons (the hypertext links) or keys. The entry-level page at a particular node is commonly referred to as that node's {\bf front page}. On platforms which allow hypertext-processing (e.g., mouse-driven systems), the user "points and clicks" on these buttons in order to follow a link. Figure~\ref{Midas} illustrates how one of these pages may be represented using the {\bf Midas} browser. Each of the text entities enclosed in a rectangle on this page is a link.

```
\begin{figure}[t]
\psfig{figure=midas.ps}
\caption{xxxx}
\label{Midas}
\end{figure}
```

The default "line-mode" or TTY browser presents these links as numeric menu items. The number of the link to be followed is typed by the user on the command line. Consistent with WWW's hypertext metaphor, none of these buttons/links need be sequential or hierarchical in their relationship to one another.

Figure~\ref{Index} illustrates how a link can be a dynamic request to an index server. Keywords for index searches are usually typed in a window especially provided for that purpose. Figure~\ref{MidasIndex} shows the Midas interface for keyword searches. In this particular example, the index server is the WAIS gateway providing access to the {\it infamous} CIA World Fact Book. The keyword to be used for full text searching (e.g., the name of a country), is entered in the keyword box which is now active.

```
\begin{figure}[t]
\psfig{figure=Index.ps}
\caption{xxxx}
\label{Index}
\end{figure}

\begin{figure}[t]
\psfig{figure=midasindex.ps}
\caption{xxxx}
\label{MidasIndex}
\end{figure}
```

Consistent with WWW's distributed information model, the links followed from a page may be on many different machines anywhere on the network. As the user pushes buttons and browses one file after another, he/she may be actually querying information servers throughout the network, exploring therefore a large information {\bf web}. The traversal of this web is usually transparent to the user in that WWW handles all of the necessary network protocols.

\subsection{General Features of WWW}

The general features of WWW may be summarized as follows: [Cailliau et al92]:

\begin{itemize}

\item{WWW Can Insure Access to Most Current Information}

Information need only to be represented once, as a reference to the original/master source document may be made instead of making a local copy.

Links allow the topology of the information to evolve, so modeling the current state of a topic area of interest is possible at any time without constraint.

\item{WWW Optimizes the Use of an Information Source}

The user-available information web stretches seamlessly from small personal notes on local workstations to large databases anywhere on a network.

Indices are documents, and so may themselves be found by searches and/or following links. An index is represented to the user by a {\bf cover page} that describes the data indexed and the properties of the search engine.

The documents in the web do not have to exist as files; they can be {\bf virtual} documents generated by a server in response to a query or document name. They can therefore represent views of databases, or snapshots of changing data (e.g., time-dependent data, jointly-authored works-in-progress).

\end{itemize}

\section{WWW Conventions}

As described earlier, the success of WWW operation depends upon three conventions - a set of protocols, Universal Resource Locators and a set of data formats.

\subsection{Protocols}

A number of existing protocols, and one new protocol, form the set with which a WWW browser is equipped. Standard protocols used are:
\begin{itemize}
\item{FTP}

FTP (File Transfer Protocol) allows access to libraries of documents, software, etc. available on FTP servers which allow "anonymous" access.

\item{NNTP}

NNTP (Network News Transfer Protocol) allows access to Usenet Netnews servers and the display of news groups and articles.

\end{itemize}

A WWW browser is also equipped to handle two of the protocols of other information retrieval systems discussed in this book:

\begin{itemize}
\item{WAIS}

Allows gateway access to WAIS (Wide Area Information Servers). WAIS is described in detail elsewhere in this book.

\item{Gopher}

Allows gateway access to the Internet Gopher system. Gopher is described in more detail elsewhere in this book.

```
\end{itemize}
```

A new protocol, HTTP (Hypertext Transport Protocol), was developed to allow document retrieval and index search [Bern91]. This protocol coordinates communication between WWW browsers and servers and is easily implemented in shell script, perl, C or other programming languages. HTTP is designed to be as efficient and compact as possible so as to minimize round trip delays between the browser and server nodes. A great deal of its flexibility lies in the Universal Resource Locator. A simple data flow representing a request for a document is shown in Figure~\ref{HTTP}.

```
\begin{figure}[t]
\psfig{figure=http.ps}
\caption{xxxx}
\label{HTTP}
\end{figure}

\subsection{Universal Resource Locator}
```

The Universal Resource Locator (URL) is used to define the essential information in a link. In earlier WWW documentation, the URL was referred to as the UDI (Universal Document Identifer), but has been changed to reflect the widening variety of information resources now available. The URL is designed to be compact and printable (i.e., contain no special characters). The URL contains fields identifying the applicable protocol, a server specification (usually just the fully-qualified Internet name), a port address, a document to be identified on that server, and a search to be performed if necessary. It may optionally contain a field which specifies a particular part of a document to be selected (an {\bf anchor}) when the document is presented.

Figure~\ref{URL} shows the format of the URL and gives some examples of its use.

```
\begin{figure}[t]
\psfig{figure=url.ps}
\caption{xxxx}
\label{URL}
\end{figure}

\subsection{Data Formats}
```

For WWW the lowest common denominator for data is plain text. This format allows the ready display of simple documents and text output of index searches.

WWW servers may also generate simple hypertext documents written in a specially-designed text-processing language, {\bf HTML}. HTML (Hypertext Markup Language) is a subset of the popular SGML (Standardized General Markup Language). It has a few simple formatting options (tags) which allow it to be effectively used for on-line documentation as well as menus and search results. HTML permits browser-dependent formatting of hypertext documents and the definition of links. Figure~\ref{HTML} gives a simple example of an HTML document illustrating some of the markup tags which can be used. In this example, \#anc would contain a URL (Universal Resource Locator) to create a hypertext link.

```
\begin{figure}[t]
\psfig{figure=html.ps}
\caption{xxxx}
\label{HTML}
\end{figure}
```

As formats for representing data continually evolve, access to files of these data types must become generally available. The WWW architecture proposes a negotiation between client and server to agree on a document format for transmission.{\bf Format negotiation} is available for some formats in specific browsers. Formats of particular interest are graphic files (e.g., GIF, TIFF), text-formatted files (e.g., DVI, RTF) and sound files.

\section{How to Get More Information on WWW}

\subsection{Public Access to WWW}

A fully implemented version of WWW is available via {\bf Telnet} to {\bf info.cern.ch}. No username or password is necessary in order to use this service. This example, since it is via Telnet, uses the linemode browser. The user will be able to explore the web, but will not be able to use a hypertext-based interface.

The {\bf front page} of {\bf info.cern.ch} is highly tailored for use by CERN, the European Particle Physics Laboratory in Geneva, Switzerland (though it changes quite often). However, links do provide access to a wide variety of other information sources, including WAIS and Gopher (discussed elsewhere in this book). For example, following links entitled {\bf academic information} lead to resources in the areas of astronomy and astrophysics.

\subsection{Obtaining WWW Software}

Access to WWW server and client software for a wide variety of platforms is available from numerous sources via anonymous FTP. Using {\bf Archie} (described elsewhere in this book), a user can identify a convenient FTP site from which to get the software by issuing the command {\bf prog www} or {\bf prog WWW}. Pointers to sources for all of the popular browsers (e.g., Viola, Midas, tkWWW, XMosiac, etc.) are also available. Archie can also be used to point the user to the wealth of WWW documentation that is also available.

The guaranteed latest WWW server software (for a wide variety of platforms) and some browser software (including the linemode browser) is available via anonymous FTP from {\bf info.cern.ch}

\subsection{Additional Information Sources}

There are presently two electronic mailing lists available to those interested in WWW -

\begin{itemize}

\item{www-announce}

This list keeps subscribers informed as to WWW progress, new software releases and new data sources.

\item{www-talk}

This list provides a mechanism for technical discussion for those users interested in developing WWW software, WWW protocol evolution and WWW-related technologies.

\end{itemize}

In order to subscribe to either of these lists, send electronic mail to {\bf listserv \@ info.cern.ch}. The body of the mail should contain only the line:

{\tt add www-announce}

or

{\tt add www-talk}

The name and address occurring on the From: line of this mail will be added to the mailing list(s).

\newpage

\section{References}

Bern91
T. Berners-Lee, "HTTP As Implemented in WWW", CERN, December 1991

Berners-Lee et al92
T. Berners-Lee, R. Cailliau, J-F. Groff, B. Pollermann, *World-Wide Web: An Information Infrastructure for High-Energy Physics, Proceedings, "Software Engineering, Artificial Intelligence and Expert Systems for High Energy and Nuclear Physics," La Londe-les-Maures, France, January 1992

Cailliau et al92
T. Berners-Lee, R. Cailliau, J-F. Groff, B. Pollermann, *World-Wide Web: The Information Universe:, Electronic Networking, Meckler, Spring, 1992.

Kahn et al88
Kahn, Pau, Meyrowitz, *Guide, HyperCard, and Intermedia: A comparison of Hypertext/Hypermedia systems. (IRIS Technical Report 88-7), Brown University, Providence RI USA, 1988.

Krol92
E. Krol, The Whole Internet - Users Guide \& Catalog, O'Reilly \& Associates, 1992

Nelson90

T. Nelson, *Literary Machines 90.1, Mindful Press, Sausalito CA USA, 1990

```
\end{document}

\bye
```