

Publish a Book on the Web



A Robelle Tutorial for Novices

Bob Green, August 1996



Copyright 1996, Robelle Solutions Technology Inc.

1

For over ten years Robelle has been hosting the Seldom Met Users Group, which congregates at the Robelle booth at HP user group conferences. The *SMUG Book* is the unofficial record of these events, full of history, tips, and trivia surrounding HP computers.

In 1994, Neil Armstrong of Robelle suggested that we update and re-issue our *SMUG Book Pocket Encyclopedia*, adding information on UNIX, Windows, PA-RISC, POSIX, and networking. Not only did we update *SMUG Book*, we took advantage of the latest technology and at the same time made it available to a wider audience--we released it first on the World Wide Web.

I'm Bob Green, and in this paper I will discuss how I prepared the *SMUG Book* for its release on the Web. You can take a look at the *SMUG Book* at the following Web location:

<http://www.robelle.com/smugbook>

Qedit and Prose are trademarks of Robelle Solutions Technology Inc. All other trademarks are the property of their respective holders, which are not associated with Robelle, Qedit, or Prose.

For Techies

WWW

World Wide Web

URL

Universal Resource Locator

The URL is a unique address for a Web page. For example:
<http://www.interex.org>

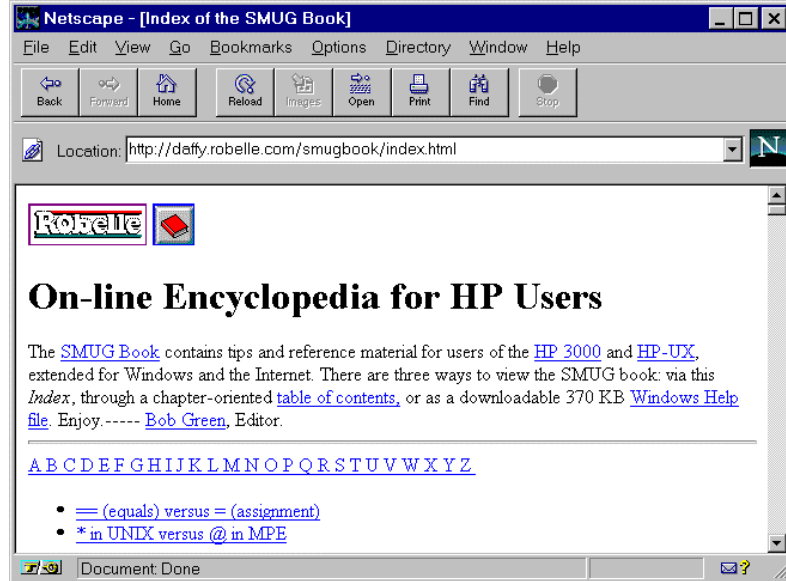
References



Best search engine for help on new topics because the editors actually review the pages they link to. If they can't find it, they can pass the search to other engines.

<http://www.yahoo.com>

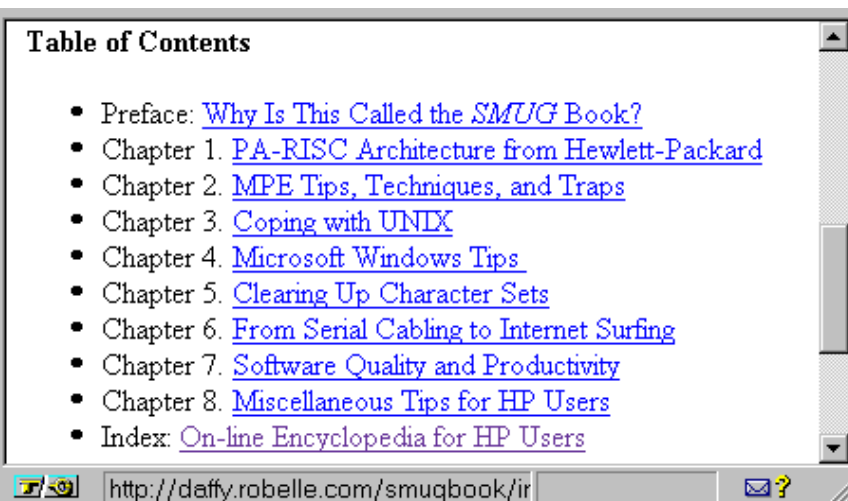
Over a Hundred Web Pages for HP Users



2

The World Wide Web version of our *SMUG Book* was converted from the Prose text formatter to raw HTML format using Robelle's Qedit, then fine-tuned and linked into pages using HTML Writer by Kris Nosack and checked for errors with Weblint. The graphics were manipulated mainly with Graphic Work Shop, with pixels edited by IconWorks and MS Paintbrush. The WinHelp file was generated using HTMHlp. I'll talk more about these tools later.

The result of using all these tools was a quick reference "book" that can be accessed in a number of ways: as a pocket encyclopedia as shown above, as a traditional book with a table of chapters as shown below, or at random by following the many links among and beyond the pages.



For Techies

HTML

HyperText Markup Language

Hard-coded pages:

The default page name in any web directory such as /smugbook is welcome.html. If that is not found, the browser looks for index.html.

References



O'Reilly publishes the best books on UNIX and the Internet topics. Several are referenced in this paper.

<http://www.ora.com>

2

2

What's Inside

- Design a Web document that is readable and Web friendly
- Get help from the Web on creating Web pages
- Learn HTML as you create your first Web page
- Check your Web pages for HTML errors
- Make your Web pages work on everyone's browser
- Optimize your pages for quick download over modems
- Convert Web pages into a single WinHelp file



3

At Robelle, we have found that O'Reilly publishes excellent books on Internet and UNIX topics. For this reason, their books are mentioned in this paper whenever a topic warrants it. We even have an O'Reilly page:

<http://www.robelle.com/o-reilly.html>

Other useful resources are listed under the "References" section on the bottom right-hand corner of these workbook pages.

There are many things to consider in Web publishing, things such page design, making pages compatible with all browsers, and optimizing document size for good performance with slow modems. "Publish a Book on the Web" aims to be a practical primer on how to put your work on the Web, based on my trials and tribulations while I prepared the *SMUG Book* for Web release.

I assume you know a few basic things about the World Wide Web (WWW). For an excellent introduction to the Web, to browsers and to servers, go to David Greer's paper "Client/Server, the Internet and WWW":

<http://www.robelle.com/www-paper/overview.html>

For Techies

HTTP

HyperText
Transfer
Protocol

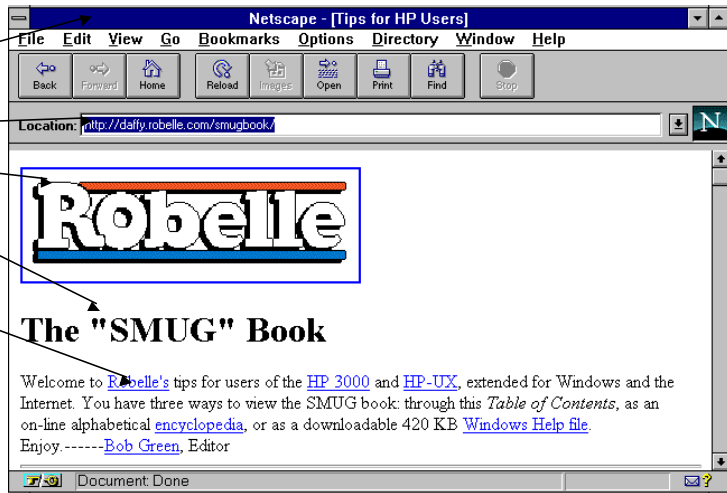
References

Come to us for HP 3000 and HP-UX information, including tutorials and WinHelp files you can download:

<http://www.robelle.com>

Components of a Web Page

- <title>
- URL
- In-line Graphic
- Heading <H1>
- Links

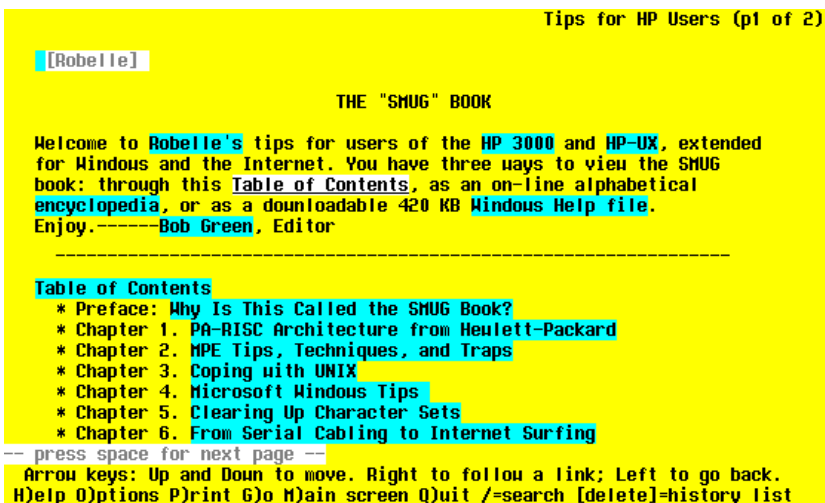


4

You experience Web pages through a browser such as Netscape or Mosaic. I don't think I have to explain how this works:

You click, therefore you exist!

Different browsers may display the same page in different ways. There is even a non-graphical browser called Lynx for those users with only a terminal interface to the Web:



For Techies

<title> and <h1> are HTML tags. Tags define how a page should appear logically, but not physically. The browser determines what physical fonts and colors are used, based on what is available on the client.

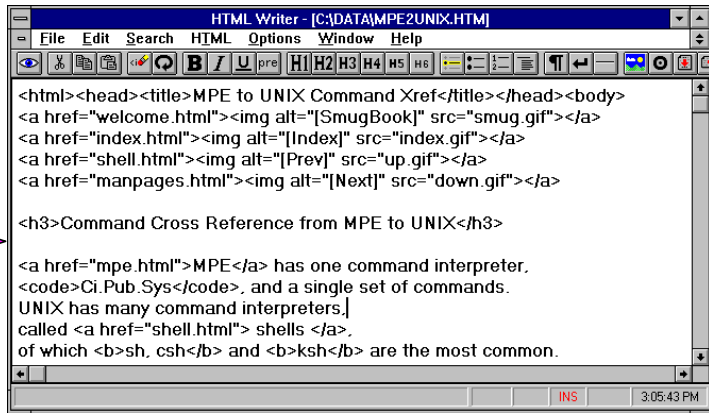
References



HTML: The Definitive Guide. April 1996. O'Reilly. "How to create beautiful, informative Web documents you'll be proud to display."

HyperText Markup Language

- Embedded formatting tags
- Like TDP, troff, Prose
 - `SMUG` Book becomes
 - **SMUG** Book
- **HTML Editors:**
 - HTML Writer
 - HotMetal - WYSIWYG
 - HotDog - all tags
 - Netscape Gold 2.0



There are as many ways to edit HTML files as there are Web authors. At a recent meeting of 42 programmers, some were using HTML Assistant for Microsoft Word, some were using Qedit or vi to edit the HTML directly, and one was using HotMetal (a fancy WYSIWYG editor).

I use HTML Writer because it restricts me to HTML 2.0 (but still allows me to Underline). When I let HTML Writer insert the tags, I can be confident that my page will work on all browsers. I keep a copy of Netscape running (as well as one of Mosaic if I am experimenting with a new HTML tag), and switch back and forth between editor and browser every 10 minutes. I edit the page with HTML Writer, then do Save File, Alt-Tab to get to Netscape, and Alt-View-Reload to see the result. Although there does not seem to be any development happening for HTML Writer, it has only a couple of minor bugs, and the Help file is excellent. You can get it off the Net for free at

<http://lal.cs.byu.edu/people/nosack/index.html>

Also recommended by many people:

HTML Assistant for Word

Hot Dog Professional from Sausage Software

<http://www.sausage.com> (\$100 to register)

For Techies

HTML tags come in pairs: `<h3>` above starts a level-3 heading which ends at `</h3>`. A closing tag is usually the same as the opening, but with a slash “/” inserted. A few tags are unpaired: `<p>` paragraph, `
` line break, `<hr>` horizontal line.

References

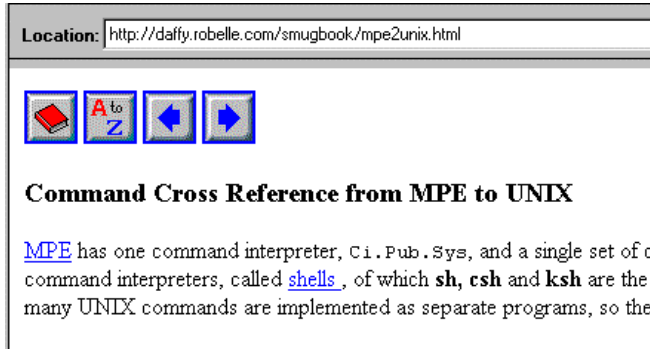


Search for:

html editors

Make Your Own Buttons Don't Count on the Browser "Back" Button

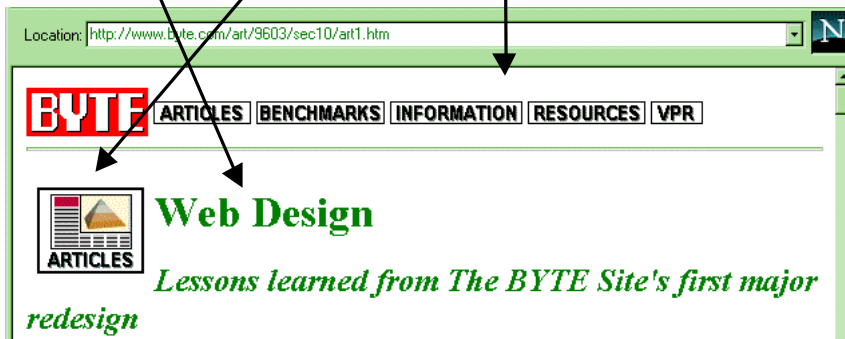
- People don't always arrive at a page the way you expect
- Show the context in first sentence, with links
- Use navigating buttons
- Be consistent: if it looks like a button, it should be a button.



6

Users may not arrive at a page by the path you expect them to use. They may jump in from a search engine with no context and, as a result, their browser "Back" button will take them back to the search engine. Users also pass favorite sites around by e-mailing URLs (site addresses) and creating links. So, remember to always put keywords in the title, heading and first sentence, and link keywords to explanatory pages. You should also add navigating buttons on every page. Another version of this idea is described in the article "Web Design" in the March 1996 issue of *BYTE* Magazine:

"We map The Byte Site with a standard header made of a row of buttons that link to the site's main areas. Below that we include the page's title, preceded by an icon that matches one of the buttons. The row of buttons is a standard device that says 'You can go to these places.' The matching title adds an important fact that many sites omit. It says 'You are here.'"



For Techies

Location:

```
file:///c:/
```

This is a URL to a file on the C drive of your PC. "|" is used instead of ":" and the UNIX-style "/" is used to separate directories instead of the DOS-style "\".

References

YAHOO!

Search for:

html bad style

How Many Browsers Do You Support?

- Tricks which mix up HTML tags in non-standard ways may work on some browsers, but will probably get you in trouble eventually
- Most browsers now handle forms
- Newer graphical browsers handle tables
- `<u>` for Underline and `<blink>` are not standard HTML. Some browsers will die when they see them. Avoid them.
- Frames and JAVA™ applets are okay--ignored by older browsers



7

Any fancy effects that you use, especially if non-standard or brand new, should degrade gracefully. For example, if you specify the colors for your page, most older browsers will ignore the HTML tags if they can't figure out what to do with them. However, if you build your page around Tables, then you have written off all the Lynx and Mosaic 1.x users. They may be a small group, but they still exist. Remember, some of the people who use Lynx do so because they are blind and need the non-graphical interface to access the information.

Do you turn off the Toolbar and Location windows of your browser to give yourself a bigger page window? Do you then design helpful frames with graphic logos in them to fit exactly within that page area? Very clever. What happens when an ordinary Netscape user who has the Toolbar and Location windows enabled accesses your site? This user gets scroll bars on the frames containing your slightly too large logos. The Web pages designed by people with 21" monitors are even worse! Probably there are Netscape HTML options to scale the graphics to fit in the client frame, but you get the idea. *Using advanced formatting options well is not as simple as it looks.*

Before you design, think about how your page will look on different monitors. Exotic wallpaper and font colors are dangerous. On many screens, such as my laptop, you lose contrast and can't read the text.

For Techies

HTML list tags are frustrating. They aren't flexible enough to do exactly what you want, which is often a table or a spreadsheet. Before you post a tricky list putting `<dt>` tags inside `<menu>` tags, check it with Weblint.

References

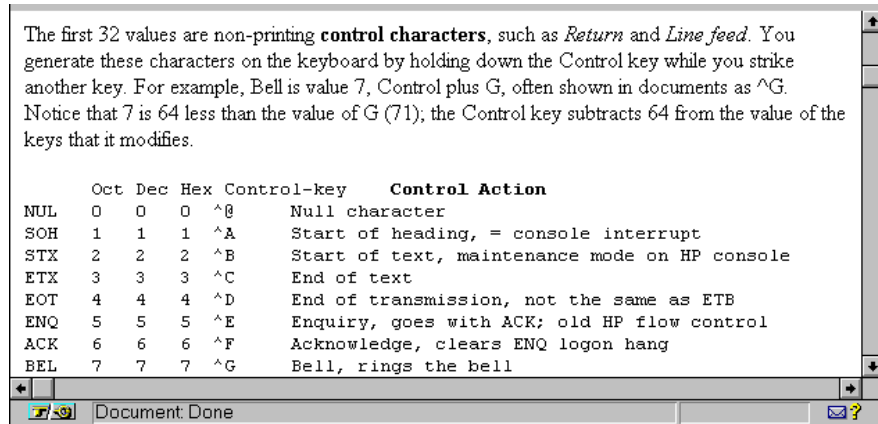
YAHOO!

Search for:

browsers,
html checkers

Getting Text to Line Up in Columns

- Tables are one option, but `<pre>` is a portable alternative
- ` ` is not an option (non-blank space)



8

Instead of using Tables, I use the `<pre>` tag to define an area of ASCII-formatted text (you can use Tab and Return and Space to make the “table” look whichever way you like, but I have converted most of my Tab characters into Spaces, in order to make my pages easier to convert to other formats, such as WinHelp).

```
<pre>
      Oct Dec Hex Control-key      <b>Control ...
NUL   0   0   0   ^@       Null character
SOH   1   1   1   ^A       Start of heading
STX   2   2   2   ^B       Start of text
ETX   3   3   3   ^C       End of text
EOT   4   4   4   ^D       End of transmission
ENQ   5   5   5   ^E       Enquiry, goes with ACK
ACK   6   6   6   ^F       Acknowledge, clears ENQ
BEL   7   7   7   ^G       Bell, rings the bell
</pre>
```

The Non-Blank Space (` `) in HTML is intended only to bind two words together into a single symbol that should not be broken apart at the end of the line. For example: Smug Book. Using several of them in a row is not defined in HTML as a way of spacing out text. It may work in the odd browser, but fails in most.

For Techies

How do you print an `&` character on a Web page, or a `<` character, or a `>` character? These are used to start tags.

Use a name for the character: **`&`** for `&`, **`<`** for `<` and **`>`** for `>`. Note that these names are case sensitive and must end with a semicolon.

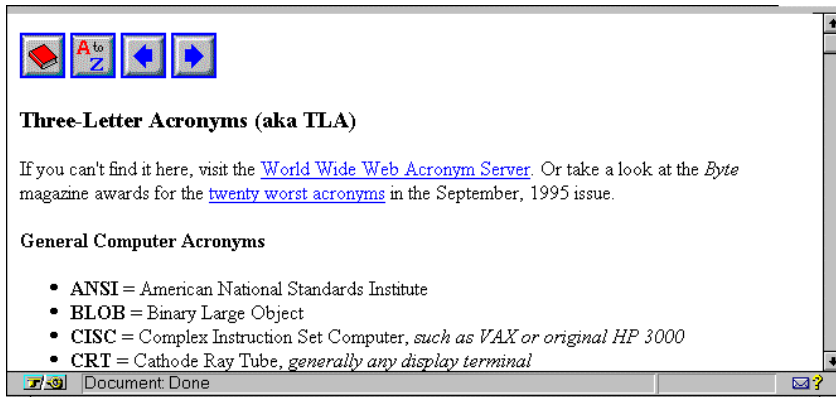
References

The *Smug Book* page on the ISO Latin characters used by the Web has several useful links.

<http://www.robelle.com/smugbook/isolat in.html>

Use Web Resources to Expand Your Scope

- Provide background by linking to tutorials and white-papers
- Create dynamic footnotes by linking to reference sources



9

In most documents you must aim at a specific audience such as “experienced HP users.” In a Web book you can make the material useful to a much wider audience by linking to other sites. If you assume that the user knows how HP VUE works, highlight *HP VUE* and link to an HP VUE tutorial page (there is one). If you can give an overview of TCP/IP, link to a site that explains it in much greater detail. The options are unlimited.

```
<h3>Three-Letter Acronyms (aka TLA)</h3>
If you can't find it here, visit the <a href=
"http://curia.ucc.ie/info/net/acronyms/acro.html">
World Wide Web Acronym Server</a>. Or take a look at
the <i>Byte</i> magazine awards for the <a href=
"http://www.byte.com/art/9509/sec7/art18.htm">
twenty worst acronyms</a> in the September, 1995
issue.

<a name="general"><h4>General Computer
Acronyms</h4></a>
<menu>
<li><b>ANSI</b> = American National Standards Inst.
<li><b>BLOB</b> = Binary Large Object
<li><b>CISC</b> = Complex Instruction Set Computer,
<i>such as VAX or original HP 3000</i>
```

For Techies

Collect feedback from your readers by including an e-mail button:

```
<a href=
"mailto:bob@gizmo.com">
<img alt=
"[mail]" src=
"mail.gif">
</a>
```



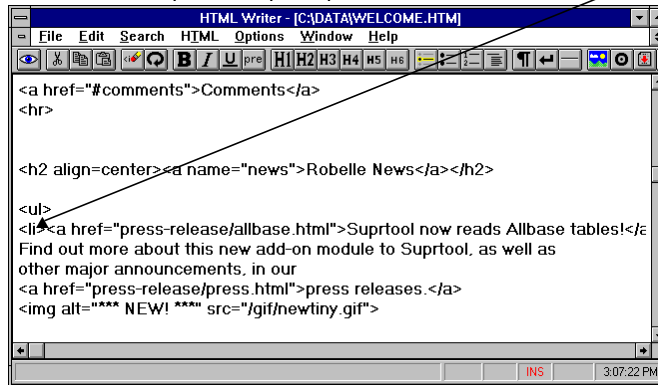
References

We have an exhaustive list of HP 3000-related Web sites, leading to many resources:

<http://www.robelle.com/hp3000.html>

Weblint Can Check Your Web Pages

- Use Perl script to check HTML files against the standard
- `$ weblint welcome.html`
- `welcome.html (36): illegal context for - must appear in <DIR|MENU|OL|UL> element.`



10

Weblint is a tool to check one or more Web pages and find violations of the HTML standard. You can pick it up off the Net at:

<http://www.khoros.unm.edu/staff/neilb/weblint.html>

There are also Web sites that will run Weblint against your pages, in addition to other sites that provide their own HTML Validation Service:

<http://www.unipress.com/weblint/>

<http://www.webtechs.com/html-val-svc/>

Weblint is written as a Perl script, which means you must have a copy of Perl on your system. Although Perl is a UNIX tool, there are versions of it for other platforms including Windows NT.

The name Perl stands for “Practical Extraction and Reporting Language.” It is an interpretive language that bridges the gap between shell programming and C programming. Perl is often used to automate system management tasks. Perl is catching on because it is a single consistent tool that works the same way on all platforms (and all flavors of UNIX). It does everything that Sed and Awk do, but you don’t have to worry about the variations from vendor to vendor. And, because it’s a single tool instead of a dozen slightly inconsistent piped tools, the coded solutions in Perl are easier to understand and run more efficiently.

For Techies

Other uses for Perl:

- pre-process Web pages to automate a consistent style
- process Web forms and return results to users
- combine many Web pages into a single HTML document
- search databases and return matches

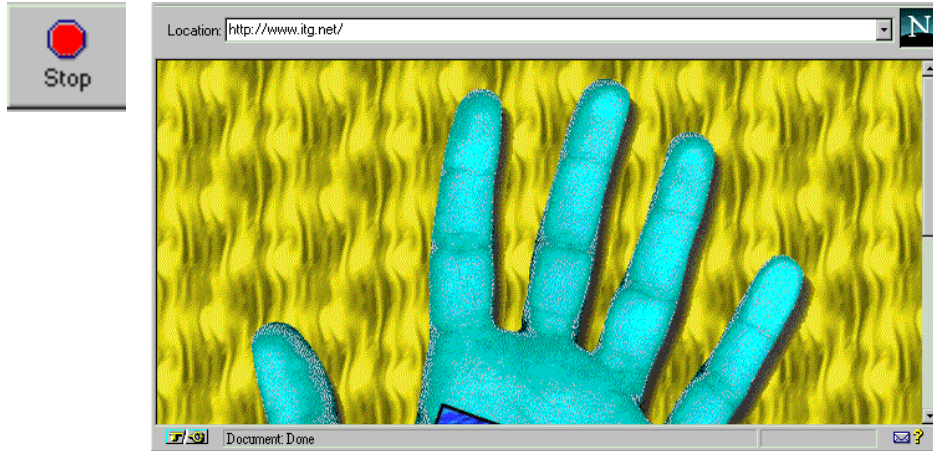
References



Wall and Schartz. *Programming Perl*. 1st Ed. January 1991. \$29.95. “Authoritative guide to the hottest new UNIX utility in years, co-authored by its creator, Larry Wall.”

How Long Does Your Page Take to Load?

- Users may click Stop before they see your message.



11

When publishing a long document on the Web, you must think about breaking it into smaller pages. One to 10 kilobytes (KB) is a good size. One Web page should be about the equivalent of one to three paper pages. Once you've sorted out your page size, you must then take care with the size of your graphics.

The `http://www.itg.net` page shown above has a 93KB file `hand.gif`, and a small background file `tile.gif`. But it has no text or any links to any other information!

Small GIF or JPEG files of a few hundred bytes load quickly and once cached by your browser, reload instantly. The icon GIF files in the *SMUG Book* are about 200 bytes each. To copy a GIF from a Web site, look in the HTML source to find the name, then Go To that name combined with the site URL. Use Save As to save the image on your PC. In Netscape 2.0 you can use View-Document Info to find out the URLs of the images and their size and format.

If you have a very large graphic that might be of interest, you can always put a small "thumbnail" copy of it on your page and link it to the big file. Then if the user wants to see the graphic, he or she can click on the thumbnail link.

These Web sites have useful icons and graphics for your Web pages:
<http://www.sct.gu.edu.au/~anthony/icons/>
<http://www2.ncsu.edu/bae/people/faculty/walker/hotlist/icons.html>

For Techies

```
<body  
background=  
./tile.gif  
<center>  
<img width=472  
height=573  
src=  
./hand.gif>  
</center>  
</body>
```

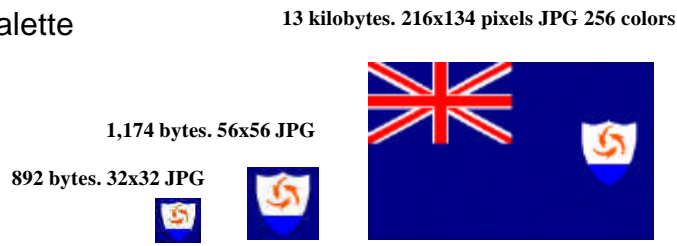
References



Designing for the Web, Getting Started in a New Medium. April 1996. O'Reilly. "Incorporating images into Web pages, resolution and improving efficiency, different browsers and platforms."

Adjust Your Graphics for Speed and Impact

- Crop out a fragment
- Scale size down
- Convert to JPEG or GIF
- Reduce color palette



12

You can reduce the size of the in-line graphic images on your Web pages significantly with a tool such as Graphic Work Shop (GWS) for Windows. Demos are available on the Net at:

<http://emoryj.jpl.nasa.gov/ftp/alchemy>

In this example, the original JPEG file used 13KB.

- I cropped the emblem out as a 56x56 file, reducing the size to 1,174 bytes.
- I scaled this back further to 32x32 pixels, which dropped the file size to 892 bytes.
- When I converted to GIF format, then reduced the color palette from 256 to 16 colors, I reduced the file again to 743 bytes. (Note: GWS often aborts when reducing the color palette. I convert to GIF first, then disable the “Quantize 15 bits” option to get it to work!)
- A person who really knew something about graphics and drawing programs could probably produce the same graphic in 200 bytes.

Most Web browsers can handle graphic files in the GIF and JPEG format, although some will accept other formats as well. GIF is the most commonly used, but JPEG often produces a more compact file for the same image. On DOS, JPEG files must have the .JPG extension. On UNIX, the extension can be either .JPG or .JPEG, either upper or lower case.

For Techies

Besides Graphic Work Shop, other useful graphics tools that you can probably find on the Net include Lview, NeoPaint, and IconWorks. Microsoft Paintbrush is useful for fixing up your pixels.

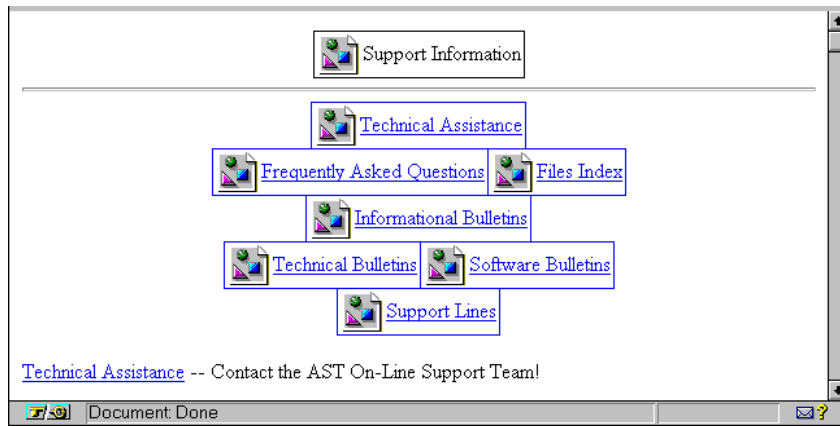
References

YAHOO!

Search for:
software graphics

The User Can Disable Auto-load Images

- Can you tell whose page this is without the company logo GIF?
- `alt="[news]"` gives alternate text -- important for link graphics



13

Users with a slow modem connection to the Internet often disable auto-load for graphical files in their browser. This makes getting to the links in the page much faster. Hopefully the user can still figure out what the links mean without the graphical images.

HTML has an option in the `` tag for this purpose: `alt=`. When the graphic is not available, the browser shows the `alt=` text instead. This eliminates the need for redundant text links after the graphic links. The page shown above has this HTML code for the first graphic image:

```
<IMG SRC="suinfo.gif" ALT="Support Information">
```

Of course this AST Support page would work better if the `alt=` tag said "AST Support Information" so you would know the company name without the company logo.

For optional "mood" graphics which do not carry specific information, you should put `alt=""` to stop some browsers from getting upset about this missing graphic.

The `alt=` tag is also very helpful to non-graphical browsers like Lynx. Because they cannot display the graphical image anyway, it gives them an alternative to display, such as `[news]` or `[search]`, and eliminates the need for a text-only version of the site.

For Techies

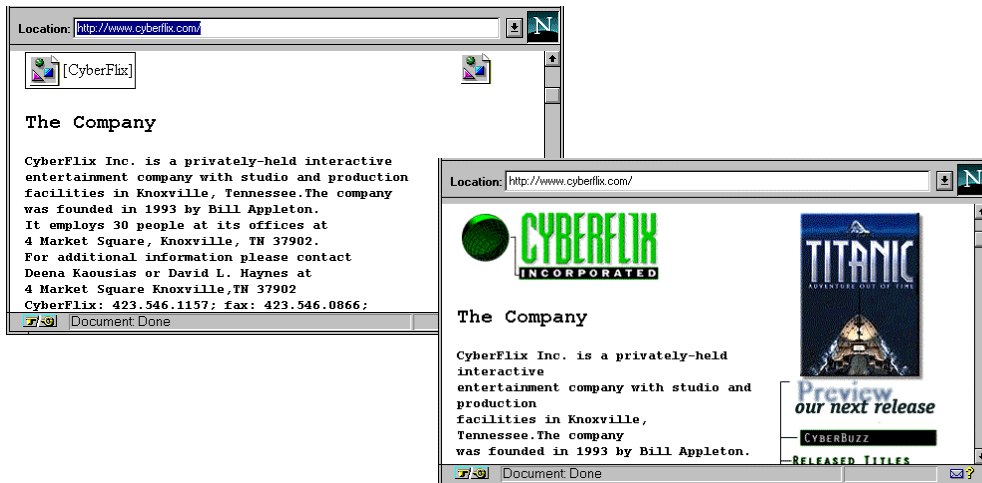
The page on this slide centers the graphics with the `<center>` and `</center>` tags around each `` tag or pair of `` tags. The dividing line is done with `<hr>`. Graphics links use `` within `<a>` anchor: ` `

References

If you prefer the pocket encyclopedia view of the *SMUG Book*, create a bookmark for

<http://www.robelle.com/smugbook/index.html>

Don't Put Everything in the Image Map



14

An *image map* is a graphical element on a Web page that contains hot-spot links hidden within it. Where you go depends upon what area of the image you click on. One problem with the image map on your page is that your user loses interactive feedback. When you move the mouse over a regular graphical link, the browser shows you the URL of the link (unless you are using JAVA to scroll irritating text through the status window at the bottom of the page). When you move the mouse around over an image map, all you get for feedback in the status area are the current coordinates of the mouse.

I do not use image maps but many sites do. However, the best sites also provide a list of text links below the image map for those users who have disabled auto-load of graphics.

Preparing an image map for inclusion on a Web page can be a daunting task. You have to select areas of the image and use the coordinates to specify a hypertext link. This can be very tedious. The Map This! freeware program from Molly Penguin Software is recommended by many people, but I haven't tried it:

<http://galadriel.eaetc.ohio-state.edu/tc/mt/>

For Techies

HTML for a Map:

```
<a href=
"http://www.xyz
.com/bin/top">
<img src=
"top.gif"
alt="Menu"
ismap>
</a>
```

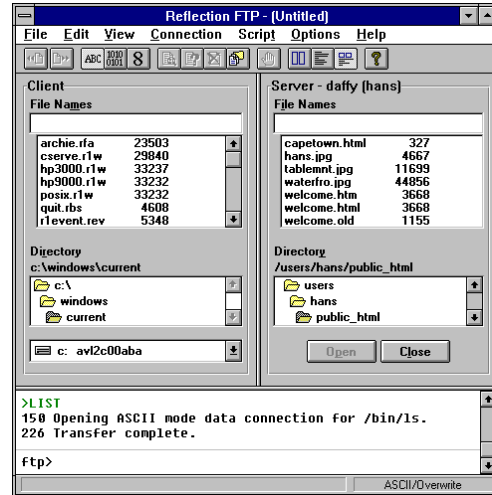
References

YAHOO!

Search for:
image maps

Installing Your Pages

- To copy your pages from a PC to a server use FTP or Reflection:
 - Use ASCII mode for HTML files
 - Use Binary mode for graphics
- Reflection can transfer multiple files (specify *.htm or *.gif)
- Use the MPUT command in FTP
- Neither changes filename extensions



15

You should create a new directory for your book in the directory for your Web site, such as `/usr/local/www/smugbook`. Put all the HTML and graphics files in it; your book's address will be `http://www.xxx.com/smugbook`. To copy Web pages from a PC to the server, you can use FTP (File Transfer Protocol) or the file transfer in Reflection's terminal emulator. You must transfer HTML text files in ASCII mode and graphic GIF/JPG files in Binary mode (aka IMAGE mode in FTP). The Reflection FTP client works very well, or you can download a Windows FTP client from the Net. I have used CuteFTP and WS_FTP with success. Try the Windows 95 site or Ultimate Winsocks site:

`http://www.windows95.com`

`http://www.tucows.com`

DOS and UNIX use different file extensions. DOS uses `.HTM` and `.JPG`, while UNIX prefers `.html` and `.jpeg`. Most browsers seem to recognize UNIX files with the DOS extensions, but it is tidier to follow the UNIX conventions. Here is a shell command from the *UNIX Power Tools* book to rename from `.htm` to `.html` after you have transferred your files:

```
%ls -d *.htm | sed 's/\(.*\)\.htm$/mv & \1.html' | sh
```

Unfortunately, when I copied back to my PC using Win95 Reflection, it retained the UNIX names in the long-filename directory and created names like `welcom~1.htm` for DOS, making all my intra-book links useless and making it difficult to run any 16-bit tools on the collection of Web pages as a whole. So perhaps the DOS extensions are better!

For Techies

To install your own home page on a system running a Web server, create the `public_html` directory in your home directory. Now copy your Web page file to `welcome.html` in that directory. If *bob* is your username at *x.com*, your URL will be:

`http://www.x.com/~bob`

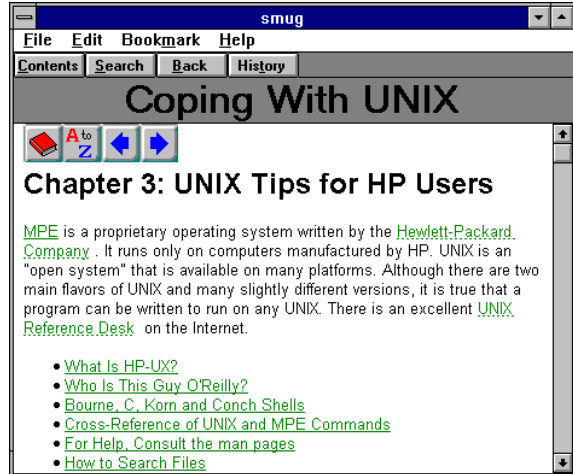
References



UNIX Power Tools.
 March 1993. O'Reilly.
 CD-ROM included.
 "Tips, tricks, concepts,
 and freeware." Best book
 for learning shell spells.

Convert Web Pages into a Single WinHelp

- Provide a single file to download
- Convert HTML to WinHelp
- Connect WinHelp to browser for external Web links
- Package as an auto-install EXE



16

For most users, the Internet is not fast enough or convenient enough for information that they access often. When they find a Web reference that would be useful every day, they want a copy of their own: either a printed copy or a local copy on their PC. They could save or print every page individually, but what they really want is a single file they can download.

We chose the WinHelp format for the *SMUG Book*. This maintains most of the web-like qualities of the pages, such as links and fonts and graphics, but now combines them into a single file. We found HTMhelp from HyperAct to convert a group of related Web pages into a single WinHelp file. Point HTMhelp at the starting point (`welcome.html`) of the book, provide a name and title for the WinHelp file, and click Convert. It finds all the linked pages on the local system, as well as the graphics files, and merges them together, going through an RTF file to a WinHelp file:

```
http://www.hyperact.com/hcata11.html
```

HyperAct has another product, WWHelp, which works with HTMhelp to make the external Web links in your WinHelp file alive. When you click on an external link, it activates your browser and passes the URL to it for retrieval. Without WWHelp, clicking on an external link causes a message window to pop up showing the URL. A third HyperAct product, HLPDK allows us to convert the WinHelp file into an RTF file for MS Word, so that we can print the *SMUG Book* too.

To turn the WinHelp file into an auto-install file, we used Setup Pro from Eschalon Development: <http://www.eschalon.com/>

For Techies

When you create many inter-linked Web pages, use relative links, not absolute. `` works when you move the files around, while `` does not. If you put all the files in the same directory, it is also easier to convert to WinHelp.

References

YAHOO!

Search for:

html converters

Summary: How to Publish a Web Book

- Break the book into small “pages” for speed when uploading
- Provide the book as a single file
- Use a consistent, well-mapped page design
- Expand your audience by linking to other sites
- Go easy on the big graphics until everyone has a T1 line
- Degrade gracefully to less powerful browsers
- Download and buy a few good tools to help you



17

To assist search engines to better index your site, put <meta> tags in the <head>. Here are some from Robelle’s welcome page.

```
<META NAME="Keywords" Content="HP3000, HP 3000, Hewlett Packard, MPE, HP9000, HP 9000, HP-UX, Computer Software, HP, encyclopedia, smugbook, smug, Qedit, editor, full-screen,Suprtool, database, utility, sort, fast ">
```

```
<META NAME="Description" Content="Robelle Solutions Technology Inc. main WWW homepage">
```

```
<META HTTP-EQUIV="keywords" Content="HP3000">
```

Now that you have your Web book done, you have to let people know about it. Since you can’t go on a talk show tour (or could you?), you need to promote the site on the Net itself.

Submit the URL of your book to the search engines, such as Yahoo, Lycos, Infoseek, Alta Vista, etc. After a delay of days to weeks, they will index your pages and make the information available to searchers.

Send e-mail announcements to your friends, relatives, co-workers, and the related web-sites that you linked to in your book.

- Post a short announcement to relevant news groups.
- Post a short announcement to relevant mailing lists.

Now, sit back and enjoy the kudos for all your hard work.

For Techies

Add a search engine to your Web site for automated indexing. We use *excite* but there are many more.

<http://www.excite.com>

References

YAHOO!

Search for:

internet announce