# *inSITE* : Into the Internet

## User's Guide

**Prepare to take control...**

No Need to Refrigerate Software

**No Need to Refrigerate Software Presents:**

# *inSITE* : Into the Internet

## User's Guide

©Copyright Alexander Schonfeld, 1996. All rights reserved

No part of this manual or software may be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval systems, without the express written consent of Alexander Schonfeld

This manual could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the manual. Alexander Schonfeld may make improvements and/or changes in the product(s) described in this manual at any time.

"inSITE: Into the Internet" is ©Copyright Alexander Schonfeld, 1996. All rights reserved.

Microsoft Windows is a trademark of Microsoft Corporation.
Netscape is a trademark of Netscape Corporation.

No Need to Refrigerate Software

# inSITE: into the Internet™ License Agreement

BEFORE USING THESE PROGRAMS (inSITE), YOU SHOULD CAREFULLY READ THE FOLLOWING TERMS AND CONDITIONS. INSTALLING THE PROGRAM INDICATES YOUR ACCEPTANCE OF THE TERMS AND CONDITIONS THAT FOLLOW.

Alexander Schonfeld provides this software and licenses its use to you, the Customer. Alexander Schonfeld retains ownership of all Proprietary Software embodied herein. You acknowledge that the unauthorized distribution or use of software received from Alexander Schonfeld will cause Material Damage to Alexander Schonfeld.

You assume responsibility for the selection of programs to achieve intended results, and for the installation, use, and results obtained from the software.

Alexander Schonfeld certifies that the Internet software and documentation provided herein has been "developed exclusively at private expense."

It is understood by both parties to this agreement that Internet software depends on the correct use of the HTML language by the millions of web pages on the Internet. You must understand that errors are inevitable and that it is your responsibility to deal with recognition errors before using the results.

Alexander Schonfeld SHALL IN NO EVENT BE LIABLE FOR ANY DIRECT OR INDIRECT DAMAGES, INCLUDING PERSONAL INJURY, RESULTING FROM ERRORS IN THE FILE GATHERING PROCESS.

## License

You may:

- use the program on a single machine
- use the registration code on a single machine (if you purchase a code)
- copy the program in machine-readable form for backup purposes only, for use on the single machine
- transfer this software and license to another party only with the prior written consent of Alexander Schonfeld

The user of any end-user application incorporating any part of the programs must agree (1) not to disassemble or reverse engineer the programs or any portions thereof and (2) to comply with all U.S. export laws.

You must reproduce and include the copyright notice on any copy, modification, or portion merged into another program. You may not use, copy, modify, or transfer the program, or any copy, modification or merged portion, in whole or in part, except as expressly provided for in this license.

If you transfer the programs, or any copy, modification or merged portion of the program to another party, your license is automatically terminated. Any attempt to disassemble the object code of this software is unauthorized and thus is illegal and will result in termination of the license agreement.

You may not commit any act which would directly or indirectly violate any U.S. law, regulation, treaty, or other agreement relating to the export or re-export of this licensed program, to which the U.S. adheres or with which the U.S. complies.

## Term

The license is effective until terminated. You may terminate the license at any time by destroying the programs along with any copies, modifications, registration codes and merged portions in any form. It will also terminate upon conditions set forth elsewhere in this Agreement or if you fail to comply with any term or condition of this Agreement. You agree upon such termination to destroy the programs together with all copies, modifications and merged portions in any form.

## Restricted Rights Legend

Use, duplication, or disclosure by the Government is subject to restrictions as set forth in subparagraph ( c )(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.227-7013

Contents

*1. Before You Begin*.....

**1.1 What is inSITE?**.....

**1.2 Checking Your Package**.....

**1.3 System Requirements**.....

**1.4 Installing inSITE**.....

**1.5 A Word About Product Support**.....

**1.6 Terms You Need to Know**.....
1.6.1 What is a URL?.....
1.6.2 What is "HTML"?.....
1.6.3 What is a "file mask"?.....

**1.7 Working with inSITE**.....
1.7.1 How Does inSITE Work With the Internet?.....
1.7.2 Where to Next?.....

***2. Introducing inSITE***.....

**2.1 Starting inSITE**.....

**2.2 Who Uses inSITE?**.....

**2.3 Quick Start**.....

**2.4 Some Features**.....

***3. Searching for Files***.....

**3.1 Using inSITE with Netscape**.....

**3.2 Scanning for Files**.....

**3.3 The Recursion Depth**.....

**3.4 The Role of Masks**.....
3.4.1 Applying Masks to Files.....

**3.5 The Starting URL**.....

**3.6 Including versus Ignoring**.....

**3.7 The Links Tree**.....

***4. Saving the Information***.....

**4.1 Filtering and Selecting**.....

**4.2 Status Indicators**.....

**4.3 Getting File Sizes**.....

**4.4 Showing Full Paths**.....

**4.5 Ignoring Errors**.....

**4.6 Building Directories**.....

**4.7 The Files Found**.....

*5. Handling Local Files*.....

**5.1 Local Directories**.....

**5.2 Local Files**.....
5.2.1 Viewing Local Files.....
5.2.2 Disk Space.....

*6. Other Functions*.....

**6.1 Configuration Options**.....
6.1.1 List Files You Already Have.....
6.1.2 Prompt on Overwrite.....
6.1.3 Leave Disk-space Free.....
6.1.4 Delete Partial Transfers.....
6.1.5 Retry Connections.....
6.1.6 Time-out in Seconds.....
6.1.7 Sort Files Found List.....
6.1.8 Show File Type Icons.....
6.1.9 Sounds.....
6.1.10 Netscape Bookmarks File.....
6.1.11 Log File Options.....
6.1.12 Make Links Local.....
6.1.13 Scan Directories Not Directly Linked.....
6.1.14 Auto Number "index.html" Files.....
6.1.15 Include HTML Bookmark Links.....

6.2 **Menu Options**.....
6.2.1 Append List of URLs.....
6.2.2 Save List of URLs.....
6.2.3 Quick Launch Netscape.....
6.2.4 Show Error Window.....
6.2.5 Set Username and Password.....

*7. Trouble Shooting*.....

**7.1 FTP Sites**.....

**7.2 Downloading Files**.....

**7.3 Error Messages**.....

**7.4 Where to Get Help**.....

*8. Order Form*.....

Note: you can find Netscape behind the "Download Netscape Now" button embedded in many pages.

# Begin

## *1.1What is inSITE?*

inSITE is a program that scans URLs (sites on the Internet) for linked files and HTML-based web pages. inSITE functions much like many on-line Internet search engines. Only rather than search a pre-compiled database of information, inSITE searches the Internet and builds a record of linked information, files and locations while you watch! You can stop this search at any time and browse the links you have collected. Then you can selectively or batch save the files or HTML pages found.

## *1.2Checking Your Package*

This package should include:

- The inSITE installation disks
- The inSITE manual
- Other (optional) flyers

## *1.3System Requirements*

To run inSITE you need a copy of Netscape, obtain it free at:

http://home.netscape.com/comprod/mirror/client_download.html

You also need:

- Windows95
- 8 megs of RAM
- 3 megs of free disk space
- Netscape installed (version 2.0 or greater preferred)
- A 32bit connection to the Internet

## *1.4Installing inSITE*

Put the disk in a 3½ in drive.

Run the "setup.exe" program on the disk.

For example, select "Run…" from the "Start" menu and enter:

a:\setup

if you put the disk in the "a:" drive.

## *1.5A Word About Product Support*

No Need to Refrigerate Software provides support via email at:

inSITE will allow you to use all these protocols in your searches.

**schonfal@engr.orst.edu**

Please keep questions as specific as possible.

Bug fix versions will be provided for download (free of charge) at:

**http://www.engr.orst.edu/~schonfal/**

Your computer will stay registered for future versions. If not you may have to contact No Need to Refrigerate Software at the above email address for a new registration code. The "Register" option under the "File" menu will only be visible if the program is not registered.

## *1.6Terms You Need to Know*

### 1.6.1What is a URL?

To begin using inSITE you must first understand the concept of a “URL”. URL is an abbreviation for Uniform Resource Locator which is the "addressing system" generally used to identify the location of a file on the internet. The URL also includes information on how to access the information or site. The specific access method is called the "protocol". The files accessed can be any type of file you might find on your own computer. Some files retrieved can even have other URLs embedded in them. These files are commonly referred to as "web pages". The location part of the URL specifies the host computer, directory and subdirectory where the particular file can be found. An example URL is:

http://www.somewhere.com/adirectory/afile.txt

This can be broken down into:

protocol://[username:[password]][@]host[:port]/[directorypath/][filename]

In the above example:

protocol = http
(there is no username, password, or port specified in the above example)
host computer = www.somewhere.com
directorypath = adirectory/
filename = afile.txt

The protocols generally available on the Internet are:
**http** = Hypertext Transfer Protocol, the most common protocol in use today, used to transfer HTML documents easily:

http://www.somewhere.com/

**ftp** = File Transfer Protocol, used to access so-called FTP sites, for example:

ftp://ftp.cdrom.com/pub/demos/

**file** = used to access local files:

file:///c|/windows/

**gopher** = a simpler version of HTTP used for files and text:

It is also important to note that inSITE is NOT case sensitive when comparing these masks to files or URLs.

gopher://something.com/somewhere/

### 1.6.2What is “HTML”?

HTML stands for “Hypertext Markup Language”. A “web page” is written in the code referred to as HTML. The web page is a text file with special text command codes embedded in it. Here is an example HTML Web Page:

```
<TITLE>FTP Error</TITLE>
<H1>FTP Error</H1>
<h2>Could not login to FTP server</h2>
<PRE>We are popular, we are full, please try again later. </PRE>
<A HREF=”http://www.somewhere.com/~someone/”>Somewhere…</A>
```

The title of the page is “FTP Error”, the “H1” and “H2” delimit heading 1 and heading 2 respectively. The “HREF” is a link to a URL probably pointing to a web page (the "http" protocol is usually used with HTML based web pages when no file is specified).

### 1.6.3What is a “file mask”?

A file mask is a way of referring to a range of strings by using one string with special characters embedded in it. In inSITE, the file masks use two special characters to represent multiple characters. The “*” and “?” represent more than just their regular values. The “*” can be used to represent any number of characters or no characters. For example:

```
fred*.txt
```

matches:

```
fred12341234lksadfjasdflkhtasdf.txt
fred.txt
fredasdf.txt
```

The “?” on the other hand represents any one character. For example:

```
fred?.txt
```

matches:

```
fred1.txt
FREDz.txt
fredA.txt
```

but does NOT match:

```
fred.txt
fred12.txt
freddd1.txt
```

The "*" and "?" characters can be used as many times as needed and can also be used together.

Simply press “F1” when you have selected the function you are interested in knowing more about. The help window will popup.

In inSITE it is also easy to combine masks. Simply put a “;” (semicolon) between masks, for example:

```
fred*.txt;smith*.txt
```

Also, you can add some comment text that is ignored by the scanning process. Simply put anything you want ignored outside of standard parentheses “()”. For example:

```
Some people (fred*.txt;smith*.txt)
```

would match:

```
fREd.txt
Smith.txt
sMith1234123412341234asdfasdf.txt
etc.
```

As you can see “masks” are very important when it comes to specifying multiple strings quickly.

## *1.7Working with inSITE*

### 1.7.1How Does inSITE Work With the Internet?

Although you don’t necessarily see Netscape involved in inSITE’s operation, Netscape provides inSITE’s interface to the Internet by returning requested information as a web page. This web page has links imbedded in it. All the above protocols return web pages that Netscape generates. These pages are in HTML (Hypertext Markup Language) format. HTML documents have URL links embedded in them at various points. An FTP site directory listing is actually returned as an HTML document with links to all the files on the FTP site. InSITE breaks these links down into two groups: HTML documents and other files. Netscape is also used to view most of the information inSITE returns to your computer.

### 1.7.2Where to Next?

InSITE is fairly self explanatory so you might be able to figure it out simply by reading the “hint bar” at the bottom of the window. Reading the manual will make things clearer if you do not have experience with the WWW (World Wide Web) already. If you have used Netscape extensively and browsed the source of a few web pages then you are probably ready to give inSITE a test run. The context sensitive on-line help is available for some extra information on specific functions.

By configuring Netscape you can view almost any type of file. Sound files, Zipped archives, movies and many other types can be viewed with Netscape if the associated utility is properly configured under Netscape.

## *1.8Starting inSITE*

InSITE should have been added to your “Start” menu. Simply select the inSITE icon from the “Start” menu and it will execute. If you do not have Netscape installed a warning message will be displayed. Netscape does not need to be running. You should have your connection to the Internet running (if required).

## *1.9Who Uses inSITE?*

Almost anyone who wants to easily harness the information power of the Internet can find a use for inSITE. Everyday users may want to capture a large collection of files while they sleep (over dial-up lines). Lawyers may want to scan whole sites for information that may have been copyrighted. Security people may want to scan their own site to see if some files they didn’t know about are available to the public. Web Masters may want to scan their sites to verify links are functioning and accessible. On the whole anyone who wants a clear file based picture of the Internet could benefit from inSITE. People who know where the information they want is, but want to get it all, will also benefit from inSITE.

## *1.10Quick Start*

When you find a site you are interested in through a news article or web page etc. you then want to explore it. Netscape is the best way to do this. When you find that the site has many files you are interested in, inSITE is becomes very useful. Run inSITE and "RightClick" on the "Starting URL" box. Select "Copy from Netscape" and the URL will be pasted into the box. Then configure the other options on the "Links" page. Once configured you press the “Scan For Files…” button and inSITE will begin scanning the pages recursively. This may take some time on a dial-up connection. Setting the recursion depth to 3 or 4 will no doubt turn up thousands of pages if not narrowed by some of the limiting factors. Entering a “Scan URL Mask” is one way of making the search narrower. For example: “*www.somewhere.com*” will only scan files with that server in the URL. Once you have scanned the pages, the files found will appear in the “Files Found” list. From there you can select which ones you want to download. There are a number of techniques you can

use to remove the ones you don't want. Before saving the files it is probably wise to check where you are putting them. See the "Save To" tab for this purpose.

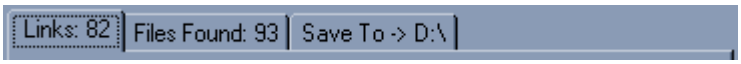
Once you have downloaded the files you can "DoubleClick" on them in the "Your Files" section of the "Save To" tab. This will request that Netscape view them.

### *1.11Some Features*

- inSITE has the ability to capture an entire site.
- InSITE can convert the links embedded in a web page to local links.
- Transfers can be logged to prove that files did exist on certain servers
- Recursion depth is limited only by how much memory you have
- Thousands of files can be downloaded without overpowering your system
- Directories not normally searched can be traced into and searched
- Local directories can be built mirroring the URLs they come from

# 4 Searching for Files

## *1.12Using inSITE with Netscape*




Normally people browse the web one page at a time clicking on links associated with pictures and graphics. These graphics and text can be arranged in any number of confusing ways and can be linked to a great number of other computers and pages with related information. InSITE strips down this massive tree of information and addresses down to its essential components. Netscape is used to retrieve the pages and thus all associated information. Everything from an FTP error message to a local file directory is converted to a web page by Netscape. Netscape provides a consistent interface to the vast and very confused network known as the Internet. InSITE takes these pages and converts them to their most basic elements: links to other files. These linked files can be basically two things: an HTML document or a file that falls into the category of "other". A number of file types (other than HTML documents) are natively displayed in Netscape which can lead to some confusion. For example, if you click on a link to a GIF file it will be displayed in Netscape. But the GIF file is not a web page! If you select "View - Document Source" you will see garbage. That is because this file is of binary nature, not ASCII text. So really, you can break everything you see in Netscape down into two groups: HTML text files or binary files and other text files. That is the only distinction inSITE tries to make. The pages which are scanned are the HTML files. Both groups can be added to the files found list.

Typically inSITE plays a secondary role. Netscape should probably be your primary browsing tool. As they say, "a picture is worth a thousand words". Thus while viewing the URL links allows you to truly see what is going on, it does not really tell you what the web pages say. The Netscape bookmark system is also fairly advanced and should be used. In inSITE you can select from the bookmarks listed by setting the location of your "bookmarks.htm" file in "Options - Configuration…". Then when you want to see the URLs in your Netscape bookmarks "RightClick" on the "Starting URL" combo box and choose "Select From Netscape Bookmarks…". A list of the URLs available in your bookmarks will appear.

## 1.13Scanning for Files

Once you have found a site of interest, you must determine the most efficient way to scan the site. This involves the use of the “Links” tab and a number of its components. The “Recursion Depth” must be set. The “Scan URL Mask” and the “Starting URL” must also be set.

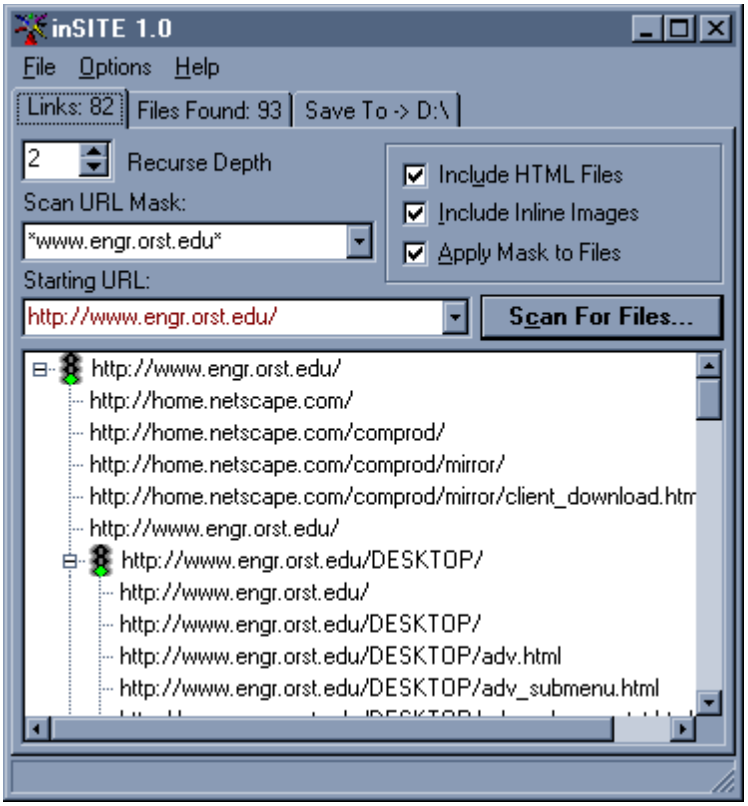



**Figure 3-1. The links scanned and available**

## 1.14The Recursion Depth

The “Recurse Depth” determines how “deep” the scan is. A “Recurse Depth” of “1” will only scan the “Starting URL”. The links found in the “Starting URL” will be added to the “Links” tree if they point to HTML files. The “Files Found” will include all the files (HTML or otherwise) linked to on the “Starting URL”. “1” is the minimum recursion depth. If you set the depth to “2” all of the links added to the “Links” tree will be scanned in turn for their links. It is easy to visualize by picturing a directory tree:

Say the “Starting URL” is “c:\junk\” with subdirectories:

```
c:\junk\morestuff\
c:\junk\morestuff\evenmore\
```

At a depth of “1”:

```
c:\junk\
```

and all the files in it would be added to the “Files Found”.

At a depth of “2”:

Links are added to the "Links" tree that are NOT scanned. These links are simply added for reference points, and future scanning. Nothing that does not match the "Scan URL Mask" is scanned. This mask is NOT case sensitive so "HTTP://*" is the same as "http://*".

```
c:\junk\
```

and

```
c:\junk\morestuff\
```

and all the files in both of them would be added to the "Files Found". At a depth of "3" all the directories and all their included files would be added to the list of "Files Found". Typically a depth of 2 or 3 is the maximum you would want to go to. When scanning a specific site as determined by the "Scan URL Mask," it is usually safe to set the depth much higher. Since inSITE will not scan the same page twice and you are sure the site is not too large, it is safe to set the depth as high as 10+. This will make sure all files are scanned.

## *1.15The Role of Masks*

The "Scan URL Mask" is key to narrowing your search. A depth of 3 on a relatively modestly connected site may easily branch out into hundreds of links. Imagine any link to a major links collection spanning thousands of pages. The solution is to use a mask. The mask lets you specify a wide range of files while at the same time narrowing your search. For more detail on what masks are, see "The Terms You Need to Know" section. Commonly you want to restrict your scan to a single server or maybe just two servers. To restrict the scan to just two servers, say "www.server.com" and "ftp.server.com", you could enter the following masks:

```
*server.com*
*tp://???.server.com/*
*???.server.com*
etc.
```

As you can see there are many ways to do the same thing with masks. I would recommend the shorter "*server.com*" technique. Sometimes you may want to limit your search to one type of protocol, say "gopher". To do this you would simply use a mask like:

```
gopher://*
```

That would effectively only allow gopher sites to be scanned.

It is important to note that the mask is applied to the FULL URL path, not the one that occurs in the web page source file.

### 1.15.1Applying Masks to Files

Sometimes it becomes important to only add files that exist on a certain server to the “Files Found” list. Pages can link to files not on their own server. This is what makes the WWW possible. However, the “Scan URL Mask” normally only determines what is scanned, not what is added to the list of files found. Thus if you only wanted the “Files Found” to be from a single server you would have to check the “Apply Mask to Files” check box. This is true because every file that is on the server includes the server name in its fully expanded URL. For example:

http://www.server.com/fred.txt

That file would match the “*www.server.com*” mask and would thus be added to the list. But, “http://www.somewhereelse.com/fred.txt” would not match the “*www.server.com*” mask and would not be added to the “Files Found” list.

## *1.16The Starting URL*

The “Starting URL” determines which page is scanned first. You can access more options for this button by “RightClick”ing on the combo box:

**Figure 3-2. Starting URL pop-up menu**

- **Copy URL from Netscape**: This option simply copies the URL currently in Netscape to the Starting URL text box. This assumes you have a Netscape window open and a URL is present in the URL selection combo box.
- **Open URL in Netscape**: This option loads the Starting URL in Netscape, this also assumes you have a Netscape window open and ready. If not, Netscape is started .
- **Add to Bookmarks**: This option adds the current Starting URL to the Netscape bookmarks file as defined in the Configuration menu. This function basically modifies the “bookmarks.htm” file to reflect the addition of the new bookmark. However, you must answer “Yes” if Netscape detects the file was modified and asks if it should be “reloaded”.

- **Select from Netscape Bookmarks**: This option pops-up a listing of the URLs in the Netscape bookmarks file as set in the “Options - Configuration” menu.
- **Select from Session List**: This option pops-up a list of all pages found in the last scan.

Like other combo boxes in inSITE the “Starting URL” box stores a certain number of URLs in its history. This number is specified in the “inSITE.ini” file in your Windows directory. The heading is “maxURLs” or something similar.

The “Starting URL” is the “root” of the “Links” tree. All links spread out from this page.

When attempting to capture a whole server it is important to find a root page that does indeed link to all other pages. Sometimes servers will not link to every available page. Usually the extra directory scanning functions will find the other pages unless the access to directories is restricted.

## *1.17Including versus Ignoring*

As with the “Apply Mask to Files” check box the other “Include” type check boxes affect what files are added to the “Files Found” list as the pages are scanned. The “Include HTML Files” check box determines whether or not HTML files are added to the “Files Found” list.

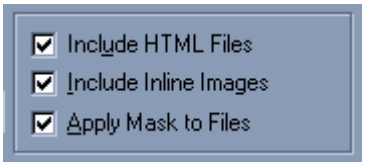



**Figure 3-3. Include and mask check group**

It is important to note that “http://www.server.com/adir/” is an HTML file. Although it may look like a directory, it does in fact return a directory listing or “index.html” file in HTML format when “saved”. That is why these files are made available in the “Files Found” list.

When downloaded a file of the “*/” type will be saved to “index.html” or “index1.html” depending on the existence of other “index.html” files. This is set in the “Options - Configuration” menu under “Auto number ‘index.html’ files”.

The “Include Inline Images” check box determines whether or not images embedded in the web page are added to the list of “Files Found”. This includes the “background” images.

This early on filtering of files is sometimes useful in narrowing the search

before it gets out of hand.

The total number of links found is listed on the top of the “Links” tab.

## *1.18The Links Tree*

The “Links Tree” is basically a tree representing all the links found in the most recent “Scan For Files”. The root of the tree is always the “Starting URL”. You can think of this “tree” as being very similar to a common “directory tree”. The difference is that these links can point back to themselves and form recursive loops. They can also point somewhere that does not necessarily point back.

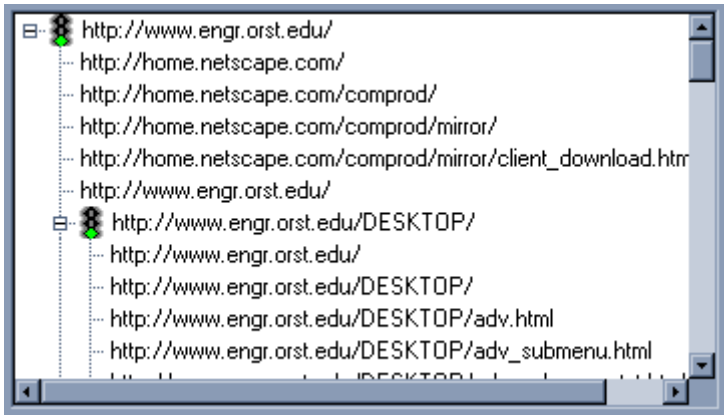



**Figure 3-4. The links available tree view**

This tree attempts to represent in an intuitive fashion the hierarchical nature of the Internet. Links that have been successfully scanned will have a “Green Light” next to them. Partially scanned links will have an “Orange Light” and failed links will have a “Red Light”. No light means that inSITE has not tried to scan that link. Thus no files will be added to the “Files Found” listing from those unscanned links.

There are a number of options not immediately visible. You can:

- “DoubleClick” on a link to **start scanning that link**. This essentially makes the link you doubleclicked on the new “Starting URL” or “root” of the tree.
- “RightClick” on the list to **sort it alphabetically**
- Press the “Delete” key to **delete a node and all sub-nodes**. This is useful for making a simpler list, but not much more.

# 5 Saving the Information

## *1.19Filtering and Selecting*

Once you have generated a list of URLs by scanning some pages you no doubt want to only save a specific number of the files. The “Save Files” button will be grayed out until some files are selected.

You can select files manually with:

- Shift+Arrow or MouseClick: This will select multiple items at once.
- Ctrl+MouseClick: Will select one additional item.
- LeftClick+Drag: Will draw a box to select more than one item

Sometimes there is such a large number of files that you need to make the list smaller. This is where the “Filter” button becomes useful. Simply enter a mask (or select from the default list) that

**Figure 4-1. The filtering and selection buttons**

identifies the type of files you want and press the “Filter” button. All files that do not match the mask will be removed.

The “Select” button uses the same mask combo box as the “Filter” button. The “Select” button, however, selects all URLs which match the mask.

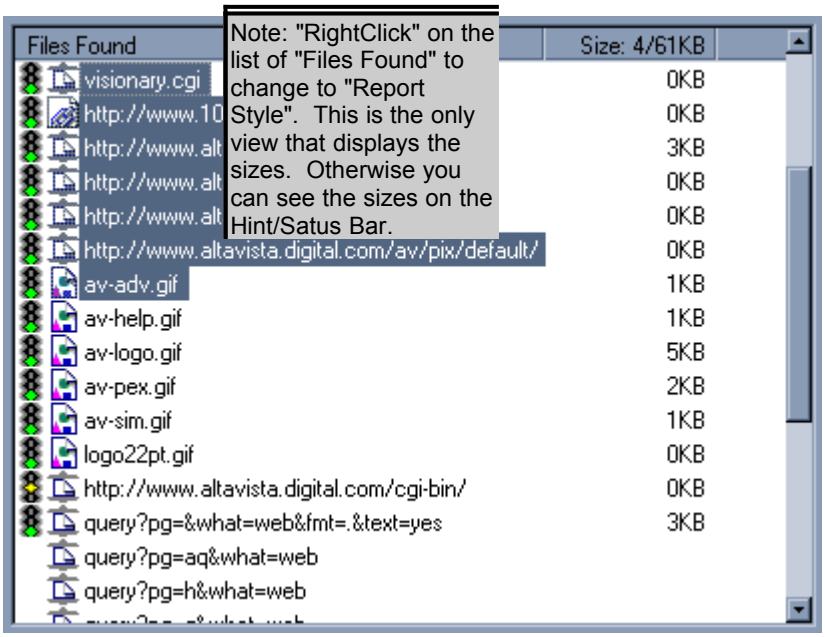
It is important to note that the “Filter” and “Select” mask is applied to the files shown in the list of “Files Found”. If the display mode is not "Full Paths" then the mask will be applied to the shortened file names. For example, “fred*.txt” would match:

fred5.txt

but would not match:

http://www.somewhere.com/fred5.txt

## 1.20Status Indicators




The "Selected" box displays how many URLs are currently selected out of the total available. The "Files Found" tab also displays the total number of available files.

**Figure 4-2. The list of files found**

The "Red Light" indicates that the transfer of the file failed. The "Orange Light" indicates that only part of the file was transfered. You can choose to have partial transfers deleted under the "Options - Configuration" menu. The "Green Light" indicates that the entire file was successfully transferred.

## 1.21Getting File Sizes

The "Get Sizes" button attempts to connect to each site and get the size of the selected URLs. The size is in kilobytes or 1000bytes. Sometimes you may want to do this more than once to verify the correctness of the sizes returned. Also some sites will not return a size for a URL. In the "Report Style" the sizes determined are displayed in the second column. The column header displays the total of all URL sizes and the total selected if any are selected.

## 1.22Showing Full Paths

The "Show Full Paths" check box determines whether or not the full URL is displayed for files which can be shortened to simple file names. For example:

> http://www.somewhere.com/something/theFile.txt

becomes:

> theFile.txt

If the URL ends in a "/" then obviously there is no short file name to use and thus the whole URL is still used. The short file name is also the one to which the file is downloaded. If no file name is present, "index.html" is used for the download name.

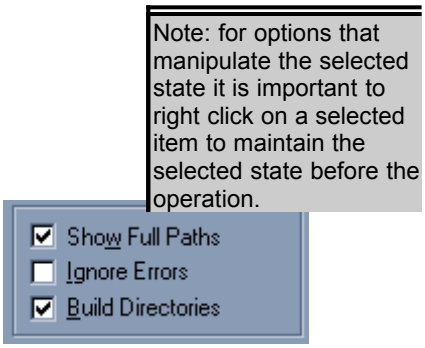



**Figure 4-3. Display and download options**

## *1.23Ignoring Errors*

This function determines whether or not inSITE will pop-up a dialog box asking for your response to an error message. It will stop the transfer and wait for your response; so on long transfers it is wise to check this box.

## *1.24Building Directories*

This determines whether or not sub-directories are built as demanded by the URL. The server name is not used and the starting directory is the one you have listed in "Save To".

For example:

> http://www.somewhere.com/something/is/here/really/farOut.txt

would build the directories:

> something\is\here\really

on top of whatever directory you listed in "Save To". If you listed "c:\junk" then it would become:

> c:\junk\something\is\here\really\farOut.txt

when you pressed "Save Files".

## *1.25The Files Found*

This list displays all the files found on the links scanned. There are a number of options not immediately visible. "RightClick" on the list to start the pop-up menu. From there you can:

- Choose the style of display:
  - **List Files**: lists the URLs as from left to right

- **Report Style**: lists the URLs as one column, and displays their sizes if known. The size column also displays the total size of selected files in the column header (the "Report Style" takes longer to display, however).

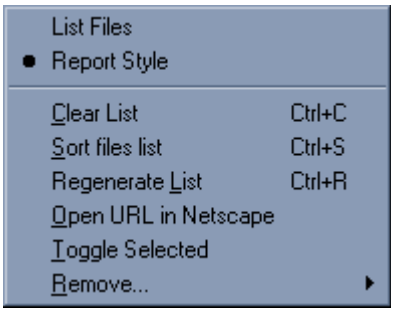



**Figure 4-4. The "Files Found" list pop-up menu**

- **Clear the list**, deleting everything.
- **Sort the list** of files alphabetically.
- **Regenerate the list**. This will rebuild a list if you feel some files are missing.
- **Open the URL in Netscape**. This will open the URL in a new Netscape window; this window will be a single bar until resized.
- **Toggle the selected files** to the opposite selection state.
- **Remove files** based on selection or download status.
- Press the “Delete” key to **delete all selected files.**

# 6Handling Local Files

## *1.26Local Directories*

This tab provides you with a directory selector. The currently opened directory is the one saved to. From the directory window you can:

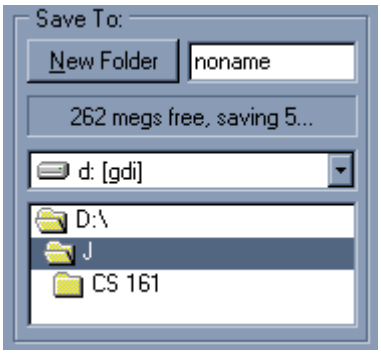



**Figure 5-1. Save to directory selection**

- Press the “Delete” key to delete an empty unopened directory

The open folder’s full path is also displayed at in the “Save To” tab.

The “New Folder” button will create a new sub-directory of the current directory. The name is specified in the text box at right of the button.

## *1.27Local Files*

The file listing at the right represents all the files in the current directory which match the file mask above. You can:

- Press the “Delete” key to delete all selected files from the directory. You can use the same keys as in the "Files Found" list to select multiple files.
- “DoubleClick” to open the local file in Netscape

## 1.27.1Viewing Local Files

**Figure 5-2. Files in open directory**

Simply “DoubleClick” on the local file in the “Save To - Your Files” box. You can configure Netscape to view almost any type of file. You can for example have Netscape launch a sound playing utility when you open a “.WAV” file. In this manner inSITE can be used to view almost any type of file. See the Netscape documentation for how to configure helper applications.

## 1.27.2Disk Space

The disk space free is displayed in the box above the current drive. If you have decided to keep a certain amount free (Options - Configuration form) that is also displayed in the box.

# 7Other Functions

## *1.28Configuration Options*

### 1.28.1List Files You Already Have

This check box determines whether or not files that already exist in the current directory are displayed in the “Files Found” list. If “Build Directories” is checked then the sub-directory that would be built is checked for the file.

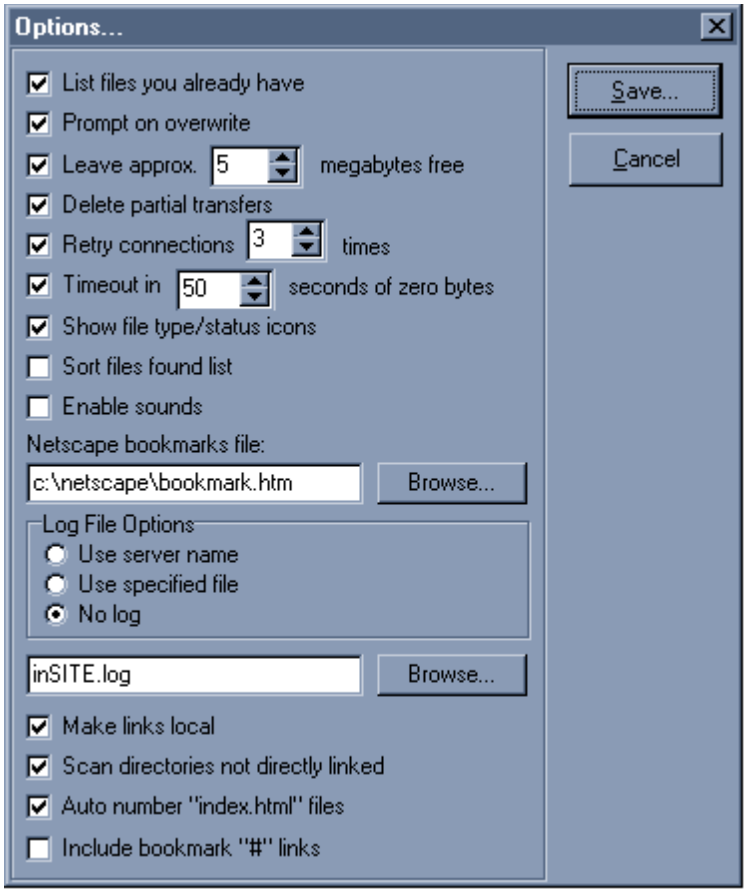



**Figure 6-1. Configation options form**

### 1.28.2Prompt on Overwrite

Determines whether or not you are prompted to overwrite files when you “Save Files” that already exist.

### 1.28.3Leave Disk-space Free

Determines if and how much disk space is saved if you are doing a large transfer. The files will not be downloaded if they exceed the amount to be kept free.

### 1.28.4Delete Partial Transfers

> Note that since FTP sites take a long time to log-in you should probably keep this value around 60 seconds.

Determines whether or not files that are not 100% transferred are saved on the disk. Unsuccessfully transferred files are usually useless unless they are large text files. Archive, executable and other binary files are typically unusable.

### 1.28.5Retry Connections

Determines if and how many times inSITE should try to re-initiate connections after failures and time-outs. You might want to set this value very high for some busy FTP sites.

### 1.28.6Time-out in Seconds

How soon should inSITE give up on a transfer.

### 1.28.7Sort Files Found List

Determines whether or not inSITE keeps the “Files Found” list sorted all the time.

### 1.28.8Show File Type Icons

Determines whether or not inSITE should assign icons based on the file extensions. This may slow some things down slightly as it takes some time to determine which file icon to use.

### 1.28.9Sounds

If you get tired of the sounds delete the ones you don't like or simply replace them with “.wav” files you do like. To enable/disable sounds, see the configuration menu. The ".wav" files are located in the "\Wavs" sub-directory of the directory you installed to.

### 1.28.10Netscape Bookmarks File

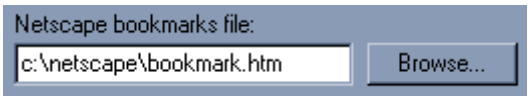



**Figure 6-2. Netscape bookmarks location**

This should be the location of your Netscape bookmarks file. This is important for the “Starting URLs” pop-up menu options. Typically the file is located under “\netscape” or “\ netscape\program” and called “bookmarks.htm” or “bookmark.htm”.

### 1.28.11Log File Options

If you wish to have a log of transfers inSITE will create a log file. The server name for this URL:

> http://www.somewhere.com/asdfsdf/asdf.txt

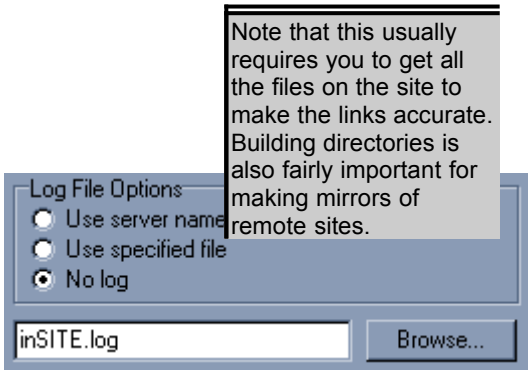



**Figure 6-3. Transfer logging options**

would generate:

> www_somewhere_com.log

The “Starting URL” is where the server comes from.

The specified file is in the text box below the “Log File Options”.

The log file stores all information about the transfers including: bytes transferred, download status, local and remote filenames and the date/time finished.

### 1.28.12Make Links Local

This attempts to convert all links in any HTML files saved to local references. For example:

> http://www.somewhere.com/frednet/something/

might become:

> file:///c|/junk/frednet/something/

### 1.28.13Scan Directories Not Directly Linked

If this is checked sub-directories not necessarily directly linked will be scanned. For example if this link was found:

> ftp://ftp.fred.com/pub/junk/misc/

Then not only that page but also:

> ftp://ftp.fred.com/pub/junk/
> ftp://ftp.fred.com/pub/
> ftp://ftp.fred.com/

would be scanned for files.

### 1.28.14Auto Number "index.html" Files

When you save a file such as “ftp://ftp.fred.com/pub/junk/” you are saving a directory listing in HTML format. It does not have a specific name so “index.html” is assigned to it. If an “index.html” file already exists then a number is appended to the file name for example:

> “index.html” might become “index1.html”

### 1.28.15Include HTML Bookmark Links

This refers to links which have a “#” sign in them. You usually don’t want to download multiple copies of a file so the bookmark information is chopped off. They can be useful for local browsing sometimes as the inSITE file list lets you launch them directly into Netscape.

## *1.29**Menu Options***

### 1.29.1Append List of URLs

This menu option lets you load a saved list of URLs and append them to the current list of "Files Found". The format of the file is (*.lks):

```
*http://www.something.com/~fred/,16KB
http://www.something.com/,
```

Where the "*" specifies whether or not the URL is selected. The file is a plain text file editable with "edit". The string after the first comma is the size of the file if it is known ("16KB" in the first example, unknown in the second).

### 1.29.2Save List of URLs

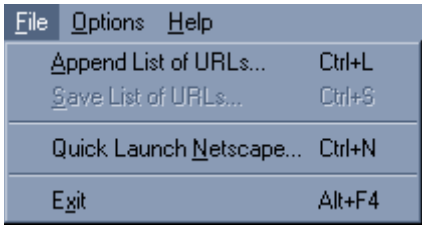



**Figure 6-4. The File menu**

This menu option saves the current list of “Files Found” to a specified “.lks” file.

### 1.29.3Quick Launch Netscape

This function quickly opens a Netscape window in compressed title-bar form.

### 1.29.4Show Error Window

**Figure 6-5. The options menu**

This checkable menu option determines whether or not the “Error Window” is open.

### 1.29.5Set Username and Password

Note: the password function only works on some sites.

This dialog box attempts to set the Netscape username and password so that you can log into protected sites.

# 8Trouble Shooting

## 1.30FTP Sites

**Figure 7-1. Warnings and error message window**

FTP sites are tricky because they take much longer to initiate a connection. Sometimes as long as 2 minutes. Also they tend to return busy signals more often than regular HTTP servers. To get around these problems you may want to increase the “Options - Configuration - Timeout in # seconds of zero bytes”. Also you may want to increase the “Retry connections # times”.

Sometimes the connection with the FTP site will be broken if you do not do anything for a specific period of time. This starts the rather long connection process once again.

If the connection appears stuck, simply restart the “Save” or “Scan” again. Since Netscape’s caching system is utilized it will go quickly until the point you were at previously is reached.

## 1.31Downloading Files

If a file appears stuck you can wait for the zero byte timeout or restart the transfer yourself.

## 1.32Error Messages

Commonly the only way for inSITE to detect an error (like “page not found - 404”) is to parse the returned web page. This is not always accurate. Since different servers can return a wide variety of “error” web pages there is very little way to be certain there was an error.

## 1.33Where to Get Help

See the email addresses listed in the "A Word About Product Support" section.

# 9Order Form

See the **ORDERFRM.TXT** file on disk.  Or select "Order Form" from the "File" menu.

## *A*

Append List of URLs · 29
Apply Mask to Files · 16, 18
Auto Number "index.html" Files · 28

## *B*

binary files · 14
bookmarks · 14, 17
Bug Fixes · 9
Building Directories · 22

## *C*

Clear the List · 23
Configuration · 26

## *D*

Delete Partial Transfers · 26
directory · 24
Disk Space · 25

## *E*

Error Messages · 30

## *F*

Features · 13
FILE · 9
file mask · 10
Files Found · 22
Filter · 20
Filtering · 20
FTP · 9
FTP Sites · 30

## *G*

Get Sizes · 21
GOPHER · 10
Green Light · 19, 21

## *H*

help · 11
HTML · 10, 11, 14

## *I*

Ignoring Errors · 22
Include HTML Bookmark Links · 28
Include HTML Files · 18
Include Inline Images · 18
index.html · 18, 22, 28
Installing · 8

## *L*

Leave Disk-space Free · 26
Links · 16
Links Tree · 19
List Files · 22
List Files You Already Have · 26
Local Directories · 24
Local Files · 24
Log File Options · 27

## *M*

Make Links Local · 28
masks · 10, 16
Menu Options · 29

## *N*

Netscape · 8, 11, 14
Netscape Bookmarks · 17
Netscape Bookmarks File · 27

## *O*

on-line help · 11
Open the URL in Netscape · 23
Open URL in Netscape · 17
Options · 26
Orange Light · 19, 21
Order Form · 32

## *P*

Password · 29
Product Support · 8
Prompt on Overwrite · 26
protocols · 9

## *Q*

Quick Launch Netscape · 29
Quick Start · 12

## *R*

Recursion Depth · 15
Red Light · 19, 21
Regenerate the List · 23
registration · 9
Remove Files · 23
Report Style · 21, 23
Requirements · 8
Retry Connections · 26

## *S*

Save List of URLs · 29
Saving · 20
Scan Directories Not Directly Linked · 28
Scan For Files · 19
Scan URL Mask · 16
Scanning · 15
Selecting · 20
Session List · 17
Show Error Window · 29
Show File Type Icons · 27
Show Full Paths · 21
Sort Files Found List · 27
Sort the List · 23
Sounds · 27
Starting · 12
Starting URL · 15, 17, 18
Status Indicators · 21
support · 8

## *T*

Terms · 9
Time-out in Seconds · 27
Toggle the Selected Files · 23
Trouble Shooting · 30

## *U*

URL · 9
Username · 29

## *V*

verify links · 12
Viewing Local Files · 25

## *W*

Web Page · 10
Where to Get Help · 31