




















-  [Introducing Internet Marauder](#)
-  [Getting Started](#)
-  [How to...](#)
-  [Custom Exploration Properties](#)
-  [Tips and Tricks](#)
-  [Registration](#)
-  [Reference](#)



-  [Introducing Internet Marauder](#)
-  [To surf is human; to maraud, divine...](#)
-  [Using Internet Marauder](#)
-  [How Internet Marauder explores](#)
-  [Getting Started](#)
-  [How to...](#)
-  [Custom Exploration Properties](#)
-  [Tips and Tricks](#)
-  [Registration](#)
-  [Reference](#)



 [Introducing Internet Marauder](#)

 [Getting Started](#)

 [Tutorial: Overview](#)

 [Tutorial: Creating the Project](#)

 [Tutorial: Setting the Project Properties](#)

 [Tutorial: Saving the Project](#)

 [Tutorial: Running the Project](#)

 [Tutorial: Viewing the Results](#)

 [Other sample projects](#)




















 [How to...](#)

 [Custom Exploration Properties](#)












 [Tips and Tricks](#)










 [Registration](#)

 [Reference](#)

-  [Introducing Internet Marauder](#)
-  [Getting Started](#)
-  [How to...](#)
  -  [Start a new project](#)
  -  [Open an existing project](#)
  -  [Choose your Marauder personality](#)
  -  [Specify file types for retrieval](#)
  -  [Use page exploration settings](#)
  -  [Exclude certain domains from a search](#)
  -  [Pause a running project](#)
  -  [Remove pages from a project](#)
  -  [Add more starting pages to a project](#)
  -  [Skip over unwanted files](#)
  -  [Determine where a retrieved file came from](#)
  -  [Access password protected sites](#)
-  [Custom Exploration Properties](#)
-  [Tips and Tricks](#)
-  [Registration](#)
-  [Reference](#)

-  [Introducing Internet Marauder](#)
-  [Getting Started](#)
-  [How to...](#)
-  [Custom Exploration Properties](#)
  -  [The Custom Properties Page](#)
  -  [Exploration Depth](#)
  -  [Image Map Exploration](#)
  -  [Retrieval Mode](#)
  -  [Tenacity](#)
  -  [Decorative Graphics](#)
-  [Tips and Tricks](#)
-  [Registration](#)
-  [Reference](#)

-  [Introducing Internet Marauder](#)
-  [Getting Started](#)
-  [How to...](#)
-  [Custom Exploration Properties](#)
-  [Tips and Tricks](#)
-  [Running multiple projects](#)
-  [Using Marco Polo effectively](#)
-  [Scheduling with command line options](#)
-  [Fine-tuning file retrieval](#)
-  [Registration](#)
-  [Reference](#)

-  [Introducing Internet Marauder](#)
-  [Getting Started](#)
-  [How to...](#)
-  [Custom Exploration Properties](#)
-  [Tips and Tricks](#)
-  [Registration](#)
  -  [Benefits of registration](#)
  -  [How to register](#)
-  [Reference](#)



[Introducing Internet Marauder](#)

[Getting Started](#)

[How to...](#)

[Custom Exploration Properties](#)

[Tips and Tricks](#)

[Registration](#)

[Reference](#)

[Glossary of terms](#)

[Copyright, Trademark, and License Information](#)

[How to contact Bourbon Street Software](#)



# To surf is human; to maraud, divine...

Congratulations! You have obtained the most powerful search and capture tool on the Internet.

Internet Marauder is a fully automated, link-following, file-retrieving web [spider](#). It will retrieve all the files you want -- and only the files you want -- from any part of the Internet. Using Internet Marauder will save you time and money, and will find files that you didn't even know existed.

You can use Internet Marauder in many ways:

- ✓ As a file retrieval tool
- ✓ As a site inventory system
- ✓ As a large-scale file search tool

Want to grab 400 sound files for a multimedia presentation? Looking for just the right background graphic? A list of every web page and file on a shareware site? Found a huge archive of great graphics -- but don't want to wait for each of them to download? Found a great link site, but don't have the time to check out every link?

No more waiting for slow pages to download. No more clicking on links for hours, only to find garbage at the end of your trail.

Use your browser for what it's made for -- browsing. Use Internet Marauder for what **it's** made for -- charting unknown web territory, retrieving thousands of files, dividing and conquering huge areas of cyberspace.

[Tell me how to use Internet Marauder](#)

[Let's get started](#)

# Using Internet Marauder

To use Internet Marauder, you create a project file that contains the address of a page on the World Wide Web. You also give Internet Marauder some rules that define what links it will follow and what files it will retrieve. You then send the spider on its mission by selecting the Start command on the File menu, or by pressing the Start button on the [toolbar](#).

Once activated, the Marauder spider will read your project's [starting page](#) and retrieve any files that it finds there. It then reads all of the links on that page, follows those links, gets the files on those pages, and so on, and so on, and so on... until it runs out of places to go.

You can tell Internet Marauder to retrieve only certain types of files, and to follow only certain types of links. For example, you may direct it to retrieve only **jpg** and **gif** files, the usual types of graphics files on the World Wide Web. You may also direct it to follow only links within the same [domain](#) as the starting page, or not to follow any links at all! Your "program" for the spider's behavior will determine how far it goes, how long it takes, and what types of files it will get.

The Marauder spider is extremely flexible! In all, there are sixteen user-customizable parameters for telling specifying which links to follow and which files to retrieve. But because this large number of parameters can daunt the new user, Internet Marauder comes pre-packaged with four of the most useful configurations, called "Explorer Personalities." Instead of setting the spider's parameters individually, simply select a personality suited to the types of pages you want to explore. Later, when you are more familiar with how Internet Marauder works, you are free to bypass the personalities and set the spider's parameters directly.

[Tell me more about Explorer Personalities](#)

[Tell me more about how Internet Marauder explores](#)

[Let's get started](#)

# How Internet Marauder Explores

Internet Marauder uses a special search and capture technique developed by Bourbon Street Software to rapidly search web pages, identify and classify their links, and then retrieve all files matching the file types you specify in the Project Properties sheet.

Internet Marauder starts at the project's first [starting page](#). The Marauder spider examines the page carefully, extracting all of its links and all of its references to embedded data.

If you have set [file type specifications](#), Internet Marauder then retrieves every file linked from that page, or embedded in it, that matches your requested file types. If you have not specified any types, Internet Marauder simply retrieves every file. Each retrieved file is stored in the project subdirectory, which is always named after your project.

*Note: The spider will retrieve linked files regardless of whether the files exist in the domain or path of the current page. In other words, the project Exploration Depth setting does not affect file retrieval.*

The Marauder spider then categorizes the page's links to other pages. If a link points to a page beyond the project's Exploration Depth, or if it points to a page in an [excluded domain](#), it is discarded. The spider then visits the remaining pages in turn; examines them for links and files; retrieves their files; categorizes their links... and so on.

If your project has more than one starting page, the Marauder spider will repeat the process above for each starting page.

As the Marauder spider explores each new page, it inserts the page's address into the [Project Map](#). You can click on any page in the Project Map to select it. The files retrieved for that page will appear in the [File List](#).

If your project directs the Marauder spider to follow links to external domains, the first link into each external domain is represented in the Project Map as a [gateway page](#). Gateway pages control the spider's access into the external domain. For this reason, other pages in the external domain are listed below the gateway page, and not below the starting page.

*Important: The Marauder spider remembers everywhere it has been!* This means it will never visit the same place twice, within the same project session. You can also prevent the Marauder spider from **ever** revisiting a page, or you can set it to revisit pages on a daily, weekly, or monthly basis. You do this by changing the [exploration settings](#) for that page.

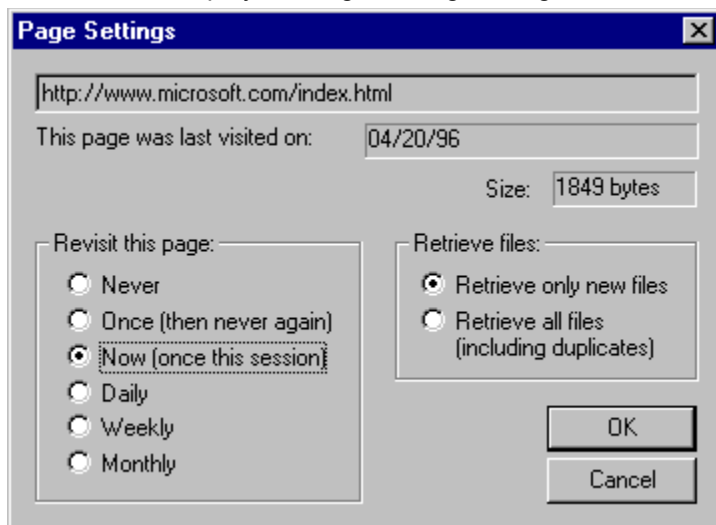
*Important: The Marauder spider remembers everything it retrieves!* This means it will never retrieve the same file twice. You can, however, force the Marauder spider to retrieve files again from a page by changing the [exploration settings](#) for that page.

# Page Exploration Settings

By default, all pages added to the [Project Map](#), including [starting pages](#) but excluding [gateway pages](#), are marked "revisit now." This means that each time you start a project session, the Marauder spider will retrieve and examine the starting page and every page below it, except those pages in [external domains](#). By default, gateway pages are marked "revisit never."

Also by default, all pages added to the Project Map are marked "retrieve only new files." This means that even if the Marauder spider revisits a page that it has already explored, it will only retrieve new files linked from that page.

You can change a page's exploration setting by double-clicking the page in the Project Map. Internet Marauder will display the Page Settings dialog box.



**Revisit Never** directs Internet Marauder to ignore this page if it is ever encountered again. By default, only gateway pages are marked "revisit never," to prevent Internet Marauder from reexploring external domains.

**Revisit Once** is a temporary setting that is convenient for directing the Marauder spider to reexplore a page only one time. After it has been explored, the page will revert to "revisit never."

**Revisit Now** is the default setting for all pages except gateways. The Marauder spider will visit pages marked "revisit now" only once, each time you press the Start button or issue the Start command from the File menu.

**Revisit Daily/Weekly/Monthly** tells Internet Marauder to revisit the page only if it has been over a day/week/month since the page was last accessed.

**Retrieve only new files** is the default setting for all pages. No matter how often it visits the page, the Marauder spider will only retrieve new files -- files that have different Internet addresses -- than any files it has previously retrieved for the project.

**Retrieve all files** directs Internet Marauder to retrieve all files linked from the page, even if it has already retrieved files with the same Internet addresses. This setting should be used only when you are sure the contents of the files have changed, even though their addresses have not. (Sometimes this is the case with "\_\_\_ of the Month" sites, which change their feature file monthly, but give the file the same address as in previous months.)

**Note:** Internet Marauder will automatically **rename** files to avoid name conflicts with files that already exist in your project folder. The program will append a "-n" to files with otherwise duplicate names, where n is an integer.

# Tutorial: Overview

This tutorial will show you step-by-step how to create and run an Internet Marauder project. The Internet site used for this tutorial hosts a collection of JPEG-encoded celebrity photographs.



**Note:** *The site used for this tutorial was active as of 04/13/96. If the site has moved or is down, running this sample project will not retrieve any files. If this occurs, try using one of the other sample projects packaged with this copy of Internet Marauder, or visit our web site at **wymple.gs.net/~bourbon** for more sample projects.*

There are five essential steps to creating and running an Internet Marauder project:

- 1 [Create a new project](#)
- 2 [Set the project properties](#)
- 3 [Save the project](#)
- 4 [Run the project](#)
- 5 [View the results](#)

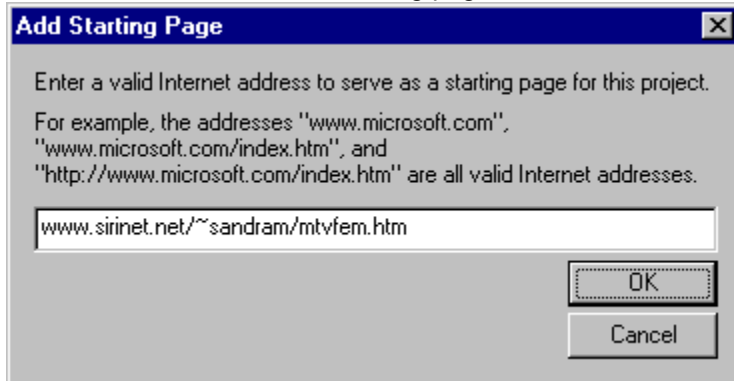
## Tutorial: Creating a New Project

To create a new project, select New Project Wizard from the File menu, or press the New Project Wizard button on the [toolbar](#). After the Wizard starts, Internet Marauder will display the Add Starting Page dialog box.

Enter the Internet address of your starting page into the Add Starting Page dialog box.



For this tutorial, the starting page is `www.sirinet.net/~sandram/mtvfem.htm`.



Note that **Internet addresses are case-sensitive**. Be sure that the address you type is exactly correct. A mistyped letter or punctuation mark will invalidate the address. Copying and pasting an address out of your browser is a good way to ensure you have a valid address. If Internet Marauder doesn't retrieve any files or follow any links for your starting page, an invalid address is the most common reason.

Press the OK button to confirm the starting page and close the Add Starting Page dialog box. You are now ready to [set your project properties](#).

# Tutorial: Setting Project Properties

If you are using the New Project Wizard, the Project Properties sheet will be opened for you automatically. If you are not using the Wizard, you must select Project Properties from the File Menu, or press the Project Properties button on the [toolbar](#). Internet Marauder will then display the Project Properties sheet.

There are four tabs on the Project Properties sheet, one for each aspect of the project: **Marauder Personality**, **File Types**, **Exclusions**, and **Custom**. The Custom tab is for advanced users; the tab is hidden until you select the Custom personality option.

At a minimum, you should set the Marauder Personality and the File Types for any new project.

Click on the [Marauder Personality](#) tab to select a Marauder Personality.



For this tutorial, set the Marauder Personality to **Geronimo**.

Click on the [File Types](#) tab to select the types of files you want to retrieve.



For this tutorial, click on the Add button and enter **jpg** into the dialog box that appears.

Click on the [Exclusions](#) tab to tell Internet Marauder where not to go.



For this tutorial, there are no changes to the default exclusions.

Click on the [Customs](#) tab to specify your exploration parameters directly, without using the pre-set Marauder Personalities. Note that you must first select the Custom Personality to enable the Custom tab.




For this tutorial, there are no custom settings.

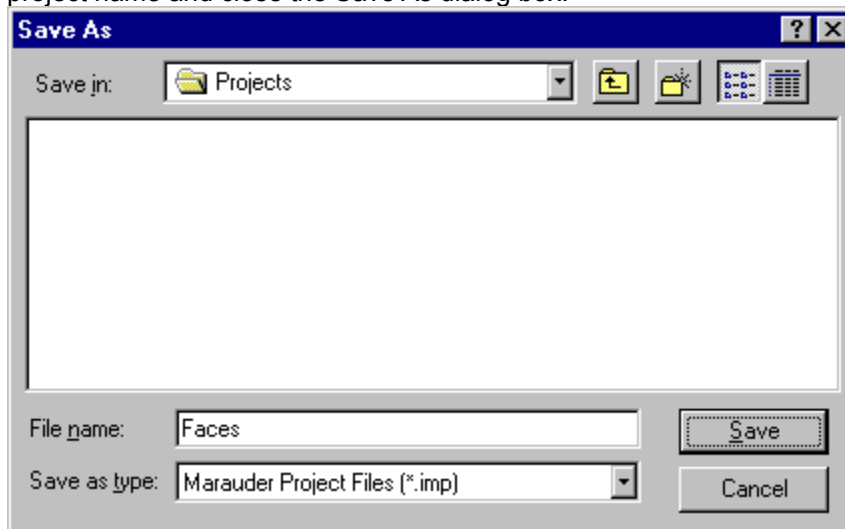
When you have finished configuring your project, click the OK button at the bottom of the properties box. You must then [save your project](#).

## Tutorial: Saving the Project

If you are using the New Project Wizard, the Save As dialog box will now be displayed for you. If you are not using the Wizard, you must select Save Project from the File Menu, or press the Save Project button on the [toolbar](#). Internet Marauder will then display the Save As dialog box.

Enter the name of your project in the File name box. You may use the list box and buttons at the top of the box to move into folder where your project will be saved.

 For this tutorial, enter **Faces** into the File name box, then click the Save button to confirm the project name and close the Save As dialog box.



Note: Internet Marauder will create a subdirectory named after your project in which to store the files it retrieves.

After you have saved the project, you may [run it](#).



## Tutorial: Running the Project

To run your project, select Start from the File menu, or press the Start button on the [toolbar](#).

Internet Marauder will connect to the Internet, request the project's first starting page, and then begin retrieving files. See [How Internet Marauder Explores](#) for more details on its search process.

At this point, you may minimize the program window, and simply let it run in the background.



**Tip:** *You might like to observe the tutorial project as it runs. The Faces project will typically retrieve about 60 files, totalling about 450 kilobytes. On a 28.8 modem, this takes about 7 minutes.*

After the project has run (and also while it's running), you can [view the results](#) in the project window.

## Tutorial: Viewing the Results

The left-hand pane of the project window is the [Project Map](#). Here, Internet Marauder shows you the pages it has explored for the project. The right hand pane of the project window is the [File List](#). Here, Internet Marauder shows you the files retrieved at each page of the project.



**Note:** *Since this tutorial project explores only the starting page (it uses the Geronimo Marauder Personality), only that page appears in the Project Map. For sample projects which explore beyond the starting page, try the [other sample projects](#) included with this copy of Internet Marauder.*

You may view retrieved files as a list of filenames, or as a report containing full file details. Use the view type buttons on the [toolbar](#) to switch between the two view types.

The retrieved files themselves are stored in the project folder, which is created by Internet Marauder the first time the project is saved, and is given the same name as the project. By default, the folder is stored in the same folder as the project file (which is, by default, in the same folder as the Internet Marauder program files).

**You can do whatever you like with the retrieved files.** You can delete them, rename them, move them, or edit them however you like. Internet Marauder creates no associations between the file names it shows in the File List and the actual files in the project folder, so there are no restrictions on what you can do with the retrieved files.

Commonly, you will want to use a viewer application to review and possibly edit your retrieved files. Examples of these are thumbnailing or graphic viewers for **gif** and **jpg** files (such as ThumbsPlus or ACDSee), the Windows 95 Media Player for **midi** sound files, and the Windows 95 sound recorder for **wav** sound files.

## Other Sample Projects

To help you get started marauding, we've included some sample projects with this copy of Internet Marauder. We encourage you to examine them, modify them, and run them, to learn more about using Internet Marauder, and how it can save you time and money when looking for files.

Open the sample projects using the Open Project command on the File menu. The sample project files are located in the "Projects" folder.

# Setting the Marauder Personality

The Marauder Personality tab of the Project Properties sheet allows you quickly to configure your project using one of four preset personalities: Geronimo, Attila the Hun, Napoleon, and Marco Polo. You may also select the Custom personality, which will allow you directly to change the search parameters for your project.

The four preset Marauder Personalities differ essentially in how far they will follow links beyond the starting page. They also differ slightly in their tenacity levels, and in how they handle decorative graphics.

**Geronimo** retrieves only files on the starting page; he does not follow any links. *Example:* For the starting page **wymple.gs.net/~bourbon/help.htm**, Geronimo will retrieve files from only the page **help.htm**.

**Attila the Hun** explores the starting page and any linked pages located in its path. *Example:* For the starting page **wymple.gs.net/~bourbon/help.htm**, Attila will follow links to any page whose address begins with **wymple.gs.net/~bourbon/**.

**Napoleon** explores the starting page and any linked pages within its [domain](#). *Example:* For the starting page **wymple.gs.net/~bourbon/help.htm**, Napoleon will follow links to any page whose address begins with **wymple.gs.net/**.

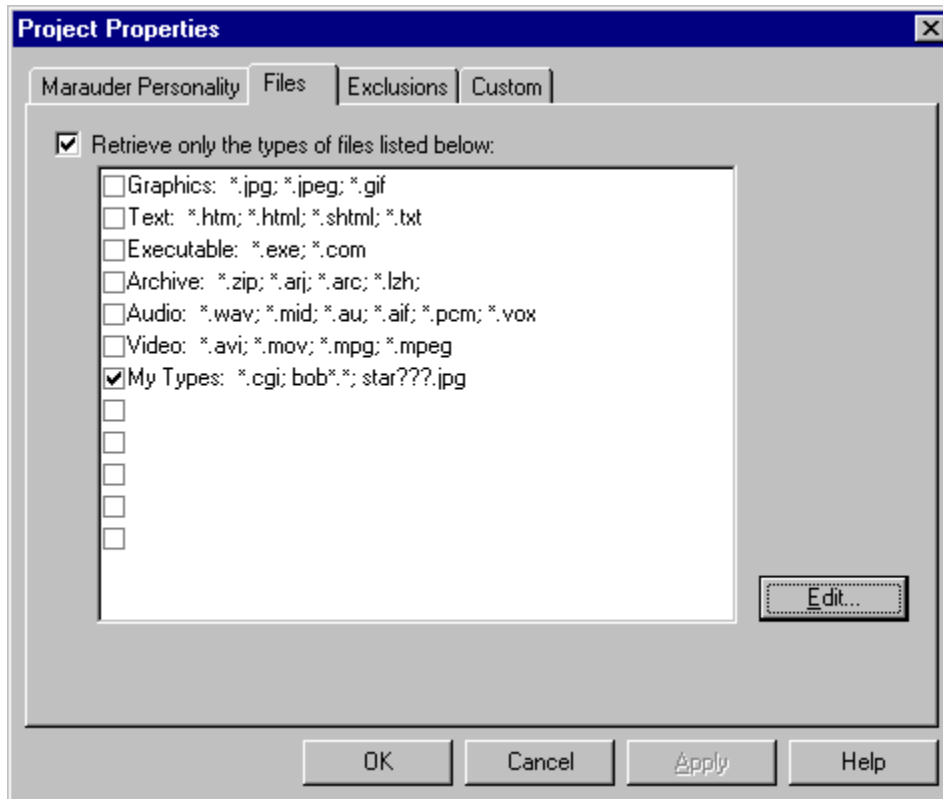
**Marco Polo** is a true explorer. He does not retrieve files; he only retrieves their names. However, he travels further than any of the other personalities. He will explore the entire domain of the starting page, and the entire domain of **any** pages linked from within that domain. Marco Polo is, therefore, the only Marauder Personality that will travel into external domains. *Example:* For the starting page **wymple.gs.net/~bourbon/help.htm**, Marco Polo will explore the entire domain of any page **linked from any page** within the domain **wymple.gs.net**.

*Tip:* You can see how a preset Marauder Personality is configured by clicking it, then clicking the Custom Personality button, and then selecting the Custom tab. Inspecting the personalities' exploration settings is a good way to learn how to program the Marauder spider.

# Setting File Types

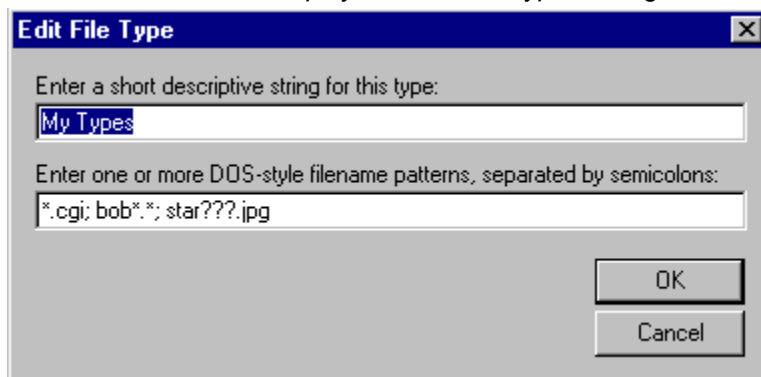
By default, Internet Marauder will retrieve every file linked from a page. You can, however, restrict file retrieval to only files of certain types, or only files that match a certain naming pattern.

Click on the Project Properties button on the [toolbar](#), or select Project Properties from the File menu, to bring up the Project Properties sheet. Select the File Types tab to display that page:



The File Types page has a pre-set list of the most common file types found on the Internet. Be sure to enable the check box marked "Retrieve only the types of files listed below," and then enable those categories of file types you wish to retrieve in the File Types check list box.

You can also create your own file type categories, by editing any of the pre-set categories or any of the blank lines in the list box. Select the category or blank line you wish to edit, then press the Edit button. Internet Marauder will display the Edit File Types dialog box:



Enter a short description for your new category, and then a list of DOS-style filename patterns separated by semicolons. There are two wildcard characters used in DOS-style filename patterns:

An **asterisk ( \* )** matches zero or more of any characters. For example, the filename pattern **\*.cgi** matches any filename having the **cgi** extension, such as **test.cgi**, but not **test.cgid..** The filename pattern **bob\*.\*** matches any filename beginning with **bob**, such as **boba.jpg** or **bob.gif** or even **bob** or **bobble**.

A **question mark ( ? )** matches any single character. For example, the filename pattern **star???.jpg** matches any filename beginning with **star** and having three more letters, and then the extension **.jpg**. This would match **star005.jpg** or **starting.jpg**, but not **star.jpg** or **starry.jpg**.

When you have finished selecting or editing file types, be sure that the check box marked "Retrieve only the types of files listed below" at the top of the file types list box is checked!

# Setting Exclusions

Besides setting a project's Exploration Depth, there are two methods for restricting where the Marauder spider will go and when it will retrieve files. **Keyword exclusions** allow you to specify a list of keywords which a web page must contain before its links are followed or its files are retrieved. **Domain exclusions** allow you to specify a list of domains that are off-limits for the spider.

To set your project's exclusions, press the Project Properties button on the [toolbar](#), or select the Project Properties command from the File menu. Internet Marauder will display the Project Properties sheet. Select the Exclusions tab.

Enter any **keywords** that pages must contain in the first text box, separated by commas. *You must check the box above the keywords list to enable keyword exclusions.* Once enabled, the Marauder spider will ignore web pages that do not contain any one of the keywords.

**Note:** *Internet Marauder will find keywords even if they are hidden inside HTML codes or comments. This means that if you are using Internet Marauder to retrieve the HTML pages themselves, you may not find your keywords in the "visible" text of the page, if you view it with your browser.*

A list of **excluded domains** appears in the checklist box at the bottom of the Exclusions page. Internet Marauder contains a default list of common domains that are usually not of interest for general file retrieval, and which are often linked from web pages. You may edit any of the default exclusion domains, and add your own to the end of the list, simply by clicking one of the list entries and pressing the **Edit** button.

# Setting Passwords

Password access is available only in our [Internet Marauder Pro™](#) product.

Besides automated password access, Internet Marauder Pro™ sports the following additional features:



Multiple retrieval threads per project -- up to 50% faster file retrieval!



Proxy server support -- Use Internet Marauder Pro through firewalls!



File size filters -- retrieve only those files that match a range of sizes that you want!



New Stealth Mode -- aggressively scans html files for links not normally visible to browsers, including commented-out links and links hidden in proprietary scripts!



Two new Marauder personalities -- Global Warlord and Guerilla Fighter!

If you are interested in obtaining Internet Marauder Pro™, please contact [Bourbon Street Software](#).



# The Custom Properties Page

Rather than use one of the preset Marauder Personalities, you can set any of the Marauder spider's exploration parameters directly using the Custom Properties page. To access this page, you must select the **Custom Personality** on the first page of the Project Properties sheet. When this personality is selected, the Custom tab will appear. Click this tab to access the Marauder spider's internal exploration parameters.

There are five groups of parameters:

- 1 [Exploration Depth](#)
- 2 [Image Map Exploration](#)
- 3 [Retrieval Mode](#)
- 4 [Tenacity](#)
- 5 [Decorative Graphics](#)

# Exploration Depth

The Exploration Depth parameter tells the Marauder spider how far away from the starting page it can follow links. Please note that Exploration Depth has **no effect** on file retrieval -- Internet Marauder will retrieve files linked from a page no matter how far away the files are located. Exploration Depth only affects the spider's link-following algorithm. It has four settings, located in the box at the upper left of the [Custom Properties Page](#).

**Explore only starting page** tells Internet Marauder to retrieve files only from the project's starting page(s). In this mode, the spider does not follow any links. Note that if your project contains more than one starting page, Internet Marauder will explore each of them.

**Stay within starting path** directs the Marauder spider to follow links only to addresses that are within the path of the project's starting page(s). The path of a page is its domain plus any local subdirectories.

*Example:* For the starting page **wymple.gs.net/~bourbon/help.htm**, Internet Marauder will follow links to any page whose address begins with **wymple.gs.net/~bourbon/**.

**Stay within starting domain** tells Internet Marauder to follow links only to addresses that are within the same [domain](#) as the project's starting page(s). *Example:* For the starting page **wymple.gs.net/~bourbon/help.htm**, Internet Marauder will follow links to any page whose address begins with **wymple.gs.net/**.

**Follow external links one level** allows Internet Marauder to go beyond the domain of the starting page. This setting is most useful for "link sites," web pages that contain lists of links to other web pages with similar content. In this mode, Internet Marauder will follow links to any page linked from within the starting domain. Once inside an external domain, Internet Marauder will change into its "stay within domain" mode and completely explore the external domain. When it has finished exploring an external domain, it will return to the starting domain and continue following links, possibly into other external domains.

***Be careful!*** When following external links the Marauder spider can travel **very far**. It can also stay out a long time, and bring back megabytes, even gigabytes of files. Unless the number of external domains linked from the starting domain is small (less than 20), projects that allow external exploration will usually require a few days to complete.

# Image Map Exploration

Internet Marauder is very adept at handling a complex World Wide Web link system: the image map. When viewed with a browser, an image map appears as a graphic image. When you click on parts of the image (often thumbnail images of a larger graphic file), the image acts as a link to a file or another web page.

Because some web pages use image maps to link insignificant or unrelated topics, Internet Marauder gives you the option of disabling its ability to follow image maps. You can access the **Explore Image Maps** switch on the [Custom Properties Page](#).

**Note:** *The image map graphic itself is a decorative graphic. Its retrieval is governed by the [Decorative Graphics](#) options, not by the **Explore Image Maps** option.*

## Server-Side Image Maps

There is a special variety of image map, called a server-side image map, that is handled by a program on the remote server computer. The server-side image map does not publish its links -- instead, it keeps them hidden, and only releases them, one at a time, when the client program (usually your browser) requests them (usually when you click on a part of the image).

Internet Marauder handles server-side image maps through a brute-force "pinging" technique. When it encounters a server-side image map, it sets up a grid across the map and "pings" the map, asking the server to release its link for each grid location. Many of these ping requests will go unanswered, as not every location on an image map will be programmed to return a link. However, with a tight enough grid, Internet Marauder can be sure of getting every available link from a server-side image map.

You can enable this advanced feature using the **Query Server-Side Image Maps** check box on the [Custom Properties Page](#). You must tell Internet Marauder what area (in pixels) it should ping over, and how far apart the grid locations should be, using the three text boxes on that page.

**Be careful!** *Pinging server-side image maps is an advanced feature. Although most ping requests take only a second or less to complete, Internet Marauder must make **many** such requests to fully explore an image map. Use grid settings below 20 only if the map graphics are very small.*

# Retrieval Mode

The **Retrieval Mode** box on the [Custom Properties Page](#) houses two unrelated but advanced retrieval options.

**Retrieve linked files** controls Internet Marauder's ability to retrieve files that are linked from a page, as opposed to those that are embedded (like decorative graphics). In most cases, you will want to leave this option enabled. If, however, you are hunting for only decorative graphics (like background graphics), you may wish to disable linked file retrieval, as it will speed up Internet Marauder's search.

**Retrieve only file names** directs Internet Marauder to skip actually retrieving files, and rather just retrieve the files' names. This option significantly speeds Internet Marauder's operation, and is handy when you wish only to get a feel for the quantity or types of files in an area. You can view the list of returned files using the [Project Map](#) and the [File List](#).

# Tenacity

The **Tenacity** box on the [Custom Properties Page](#) gives you control over how long Internet Marauder will wait for a slow server, and how many times it will request a file or page from an uncooperative server.

The **Retry connections** number box allows you to specify the number of times Internet Marauder will request a file or page from an uncooperative server. At peak access times, some servers will deny requests for files or pages on the grounds that it is too busy to respond at that time. When you are using a browser, you can sometimes "break through" by repeatedly pressing the browser's Refresh button, or simply backing up and going forward, to issue the request over and over again, until the server relents. Internet Marauder handles this automatically -- and much quicker -- by making rapidfire connection requests until it gets the file. **Note:** Internet Marauder will only make multiple connection requests if it is sure the file exists on the server. If the server responds that the file is "unavailable" (as opposed to simply "not ready at this time"), the Marauder spider will abort the request and move on.

The **Abort retrieval...** number box allows you to specify the number of seconds that Internet Marauder will wait for a response. If the server does not respond in this time, the Marauder spider will abort the request and move on. **Note:** this "timeout" parameter applies to **each** connection request; if you have told Marauder to retry connections 10 times and to wait for 60 seconds, it will wait 60 seconds for each request to be answered -- up to a total of ten minutes for a (very) slow and (very) unresponsive server. Usually, however, this does not happen, because most servers will issue a "not ready at this time" response very quickly.

# Decorative Graphics

The **Decorative Graphics** box on the [Custom Properties Page](#) gives you control over how Internet Marauder will handle graphics files that are [embedded](#) on a web page.

Embedded graphics retrieval is controlled by three switches:

**Retrieve embedded graphics** tells Internet Marauder to retrieve graphics files that are marked as embedded images, usually with the SRC (or, for advanced HTML, the DYN SRC) designation.

**Retrieve background graphics** tells Internet Marauder to retrieve graphics files that are marked as background images for the web page.

**Retrieve image map graphics** tells Internet Marauder to retrieve the graphics file used to create an image map graphic. **Note:** *This option does **not** control [Image Map Exploration](#).*

# Running Multiple Projects

Internet Marauder is fully multithreaded! This means that you can run multiple projects simultaneously. Running two or more projects at the same time can increase your overall throughput significantly, since while one project is waiting for data from a server, another project can be receiving data from another server.

All you need to do is have two or more projects open at the same time. Press the Start button (or select the Start command from the File menu) for each, and watch them all go!

Please understand that running multiple projects will slow down the progress of each project individually. However, because one project can be retrieving files while another is waiting for data, your overall throughput will increase.

***To really boost throughput, ask for Bourbon Street Software's [Internet Marauder Pro™](#)! Internet Marauder Pro™ launches multiple retrieval threads for each project -- so even single projects have an overall higher retrieval speed.***

# Using Marco Polo Effectively

Marco Polo is an unusual Marauder Personality. In fact, he doesn't really maraud -- he only explores. He is, however, extremely powerful, especially when used in conjunction with another file-retrieving Marauder Personality.

The most effective way to use of Marco Polo is to send him out to a "link site," a web page that contains a large list of links to site of interest to you. He will return a list of file names hosted on every site linked from your starting page -- the link site. You can view the files on each site using the [Project Map](#) and the [File List](#). As you select each contracted gateway page in the Project Map, you will see the list of files hosted on each gateway page's domain. If you expand the gateway page, you can browse through its domain to find the page or pages with files of interest to you.

Once you have located the domains and/or pages that have files you want to download, create a new project using one of the file-retrieving Marauder Personalities, like Geronimo, Attila, or Napoleon. **Give the new project a "dummy" starting page, then delete it using the Remove Page command.**

Return to your Marco Polo Project Map, and select a page that will be the starting page for your new project. If you will be exploring an entire domain, select its gateway page. Select Copy from the Edit menu to make a copy of that page and its files -- and, if it is a gateway page, all of the pages and files in its domain -- on the clipboard. Select your new file-retrieving project, then select Paste from the Edit menu to paste the pages into your new project.

The new project will now contain a complete map of the area you want to maraud! You must make one change before running the new project, however. Double-click each page that contains files you want to retrieve, to display its [Page Exploration Settings](#). Change the Retrieve Files option to **Retrieve all files**. This will force the Marauder spider to retrieve all of the files for that page, even though they were already visited in the Marco Polo project.

When you have finished running the new project, you may want to change each page's exploration settings back to **Retrieve only new files**, to prevent Internet Marauder from trying to retrieve those files again.



# Scheduling with Command Line Options

Internet Marauder accepts the following command-line options. You can specify these options in the command line of shortcut to the **marauder.exe** file. Separate each option with one or more spaces.

**project name:** You can specify the name of a project as a command-line parameter. When Internet Marauder starts, it will open that project automatically.

**/run** (or **/r**): This option tells Internet Marauder to begin running the project (which you must have specified by giving its name as a command-line parameter, as described above) as soon as it opens. This option is most useful for **scheduling** Internet Marauder to work on a regular basis. You can use the Windows 95 System Agent program, or any other Windows scheduling program, to activate Internet Marauder with the **/run** option.

# Fine-Tuning File Retrieval

Here are some file retrieval tips that we at Bourbon Street Software have learned from much experience in writing and using Internet Marauder:

- 1 If Attila doesn't retrieve many files from a site you **know** has lots of files, try changing to Napoleon mode, and setting [exclusion domains](#) for those links you don't want him to follow.
- 2 Turning off [decorative graphics](#) retrieval and [image map exploration](#) can significantly speed up a project when you don't care about those types of files and links.
- 3 If a page appears to have lots of [server-side image maps](#), it's worthwhile to take a look at them with your browser first. Usually they will all be the same size, and you can reduce the ping area -- and hence the time required to ping the map -- significantly.
- 4 [Running multiple projects](#) simultaneously can speed up your overall retrieval time. Just make sure that the two projects won't cover the same area, or you'll find them retrieving the same files.
- 5 Some "Item of the Week" or "Item of the Month" sites change the content of their feature file, but give it the same Internet address. Change the [Page Settings](#) for the pages that link these files to **Retrieve all files** and **Retrieve Daily/Weekly/Monthly** to ensure that you always get an up-to-date file.

## Opening an Existing Project

You can open (and even run) as many projects as you like, all simultaneously. To open a project, simply click the Open Project button on the toolbar, or select Open Project from the File menu.

Internet Marauder will display the Open Project dialog box. Select the project file you wish to open (use the file navigation list box and buttons to change folders as necessary). Click OK to confirm your choice, dismiss the dialog box, and open the project.

## Pausing a Running Project

You can **pause** a running project by pressing the Pause button on the [toolbar](#), or by selecting Pause from the File menu. The toolbar Pause button will remain depressed as long as the project is paused.

Pausing a project is useful if, for example, you are trying to download a web page with your browser while Internet Marauder is running, and Internet Marauder appears to be taking up all of your bandwidth. Pausing the project immediately suspends Internet Marauder's requests for more data, allowing your browser (or other Internet application) to communicate more freely.

When you are ready to resume running the project, simply press the Pause button again. Internet Marauder will begin running exactly where it left off.

## Removing Pages

You can remove pages from a project using the Remove Page command on the [toolbar](#), or by selecting Remove Page from the Edit menu.

There is rarely a good reason for removing a page. Removing a page effectively removes part of Internet Marauder's memory about where it has been. If it encounters a link to a removed page, it will follow that link again. Generally speaking, it is better to simply mark the page "revisit never" and leave it in the Project Map, rather than remove it.

# Adding New Starting Pages

You can add as many starting pages to a project as you like! However, please realize that every starting page is bound to the same Project Properties, so all of your starting pages will be explored in the same manner.

There are three ways to add starting pages to a project:

**Press the Add Starting Page button** on the toolbar, or select the Add Starting Page command from the Edit menu, to display a dialog box in which you can enter the Internet address of your new starting page. If you have copied the address out of your browser or Favorites file, you can simply press Ctrl+V to paste the address into this box.

**Use the Promote Page** button to promote a selected page in the Project Map to starting page status. This feature is useful if you have discovered a page within a domain that contains all the files you want. You can promote it to starting page status, remove the other pages, and change your Project Properties to match the new page's characteristics.

**Cut/copy and Paste** a page into the Project Map. You can cut or copy a page, together with its files and links to other pages, out of another project using the Cut or Copy commands on the Edit menu. When you Paste the data into another project, the first page on the clipboard becomes a new starting page. This is an effective way to use the [Marco Polo](#) Marauder Personality.

## **Skipping Files**

If you find Internet Marauder retrieving a large, unnecessary file while running a project, you can press the Skip File button on the toolbar, or select Skip File from the File menu, to abort the file retrieval process for that file. Internet Marauder will discard any data it has already retrieved for the file, and move on to the next file or page.

## Finding Files

When in List mode, the File List displays files as a multi-columned list of names, in the order in which the files were retrieved. If you want to search for a particular file in the File List, you can change the list into Report mode by clicking the Report View button on the toolbar. Once in Report mode, press any of the column headers in the report to sort the report by that column. Pressing the same column header a second time will reverse the order of the sort.

You can locate where a file came from by checking the **Page URL** column of the File List when in Report mode. The Page URL column shows the complete Internet address of the page from which Internet Marauder retrieved the file.



# Registering Internet Marauder: The Benefits

Internet Marauder is shareware. You are granted a free license to use this software for 15 days, after which the license and the software both expire.

If you decide that you like Internet Marauder – and we hope that you do! -- you can register it, keep it, and use it as long as you like. Registering the program is just like buying it.

There are really **two good reasons** to register Internet Marauder:

- 1 The program, and your license to use it, expire 15 days after you first use it.
- 2 Until you register it, the Internet Marauder will only retrieve up to 50 files per project. This may not seem like a huge limitation at first. But if you're like us – and we think that you are -- you'll get hooked on the power of downloading huge segments of the Web overnight. After a few days, you'll want to go beyond 50 files. We're confident.

And here are some other things to keep in mind about registration:



For now, registration costs only **\$10 US**. It's also about as hassle-free as we can make it. All you have to do is email us.



Your registration is valid for all “minor version” upgrades of the Internet Marauder product. This means that your registration number will work on every upgrade we release until Version 2.0. In fact, if you send us your email address when you register, we'll notify you automatically every time we issue an upgrade.

[Tell me how I can register](#)

# Registering Internet Marauder: How To Do It

To register Internet Marauder, simply send us the Registration Order Form, together with a check or credit card authorization for **\$10 US**. We accept Visa, Mastercard, and American Express.

You can print the Registration Order Form directly from the Internet Marauder program. Select Register from the Help menu to display the Registration dialog box.

Enter your name, company (optional), and at least one method by which we may contact you in the Registration Order Form text boxes. Please also select the method of payment, and include your credit card number and expiration date if you are paying by credit card. Please enter only your credit card number in the Card Number box; we can determine what type of card it is from the number.

If you are paying by credit card, you may **email**, **fax**, or **mail** the form to us. If paying by check, please only send us the form by **mail**, together with your check.

If you will **email** the form, press the **Save to Disk** button to save the order form to disk, then email us the form (either as the email message itself, or as an attached file). Our email address is

**bourbon@gs.net**

If you will **fax** the form, press the **Print Form** button print the form, and then fax it to us. If you have fax software on your computer, you can fax it directly to us from the Print dialog box. Our fax number is

**1-504-866-2937**

If you will **mail** the form, press the **Print Form** button to print the form, then mail it to us. Our mailing address is

**Bourbon Street Software  
1404 Broadway  
New Orleans, LA 70118 USA**

# Glossary

## Starting Page

The internet address at which the Marauder spider will begin its search. The starting page is entered by the user. It must be the address of a known, operational World Wide Web page. All pages linked from a starting page, either directly or indirectly, "belong" to the page and are indented below it in the Project Map.

## Custom Marauder Personality

A Marauder Personality defined by the user, using the Custom Properties page of the Project Properties sheet.

## Custom Properties Page

A page of the Project Properties sheet that is normally hidden; it appears when the user selects the Custom Personality option on the first page of the Project Properties sheet. The Custom Properties page is used to set the Marauder spider's exploration parameters directly.

## Domain

The file area contained within a server or a virtual server. The domain of the internet address "www.microsoft.com/home/software/" is "www.microsoft.com." Any internet address beginning with "www.microsoft.com" is located within that domain.

## Embedded Graphics (also inline graphics or decorative graphics)

Graphic images (and movies) that appear as part of a web page, as opposed to those that are linked from a web page (but do not appear on the page when you view it with a browser).

## External Domain

Any domain that is not the same as the domain of the starting page.

## File List

The list view showing retrieved files for the project, displayed in the right-hand pane of the project window. The File List can be switched between a list of names, and a detailed report view, using the List View/Details View buttons on the toolbar.

## Gateway Page

The first page that links into an external domain. The gateway page controls Internet Marauder's access into the external domain. By default, whenever the Marauder spider adds a new gateway page to the Project Map, the gateway page is marked "revisit never." This tells Internet Marauder to skip over that domain in later project sessions. To force Internet Marauder to reexplore the domain, change the gateway page's exploration setting.

## Image Map

An image on a web page which identifies a link or set of links, which a browser will follow when the user clicks on the image. See *also* **server-side image map**.

## Internet Address (also URL)

The full address of a page or file on the Internet. An Internet address contains the domain, directory, and name of a file. For example, the Internet address "www.microsoft.com/home/index.html" identifies the file called "index.html" which exists in the directory "home" in the domain "www.microsoft.com."

## Marauder Personality

The set of parameters that govern how Internet Marauder follows links and retrieves files. There are four preset Marauder Personalities: Geronimo, Attila the Hun, Napoleon, and Marco Polo. The user

can also define Custom Marauder Personalities.

**Page (also Web Page)**

A document, usually written in HTML (HyperText Markup Language), that may contain embedded graphics, links to other pages, and links to files.

**Path**

The directory location of a page or file. The path of the file "www.microsoft.com/home/index.html" is "www.microsoft.com/home."

**Project Properties**

Determine how Internet Marauder explores and what types of files it retrieves. A project's properties are accessed through the Project Properties sheet, which is displayed when you select Project Properties from the File menu.

**Project Map**

The tree view showing web pages that the Marauder spider has visited for the project, displayed in the left-hand pane of the project window. When you select pages in the Project Map, a list of files that were retrieved for that page will appear in the [File List](#). If the web page you select is contracted and it is a [starting page](#) or a [gateway page](#), you will see not only the files that were retrieved for that page, but also the files retrieved for any pages linked from that page.

**Server-Side Image Map**

A special kind of image map that "hides" its links from client programs, such as browsers. The server-side image map releases its links one at a time, in response to a query from a client program that identifies the location on the image that the user has clicked. Internet Marauder handles server-side image maps by rapidly making a series of such queries, each identifying a point on an imaginary grid that spans the image.

# Copyright, Trademark, and License Information

## Copyright

All software written by and copyright © 1996 Michael Del Monte.  
All Rights Reserved.

Documentation copyright © 1996 Bourbon Street Software.  
All Rights Reserved.

## Trademarks

**Internet Marauder** and **Internet Marauder Pro** are trademarks of Bourbon Street Software.

## Your End User License To Use Internet Marauder

**IMPORTANT -- READ CAREFULLY:** This End User License Agreement ("EULA") is a legal agreement between you (either an individual or a single entity) and Bourbon Street Software for the software product identified above, which includes computer software and associated media and printed materials, and this online documentation ("SOFTWARE PRODUCT"). By installing, copying, or otherwise using the SOFTWARE PRODUCT, you agree to be bound by the terms of this EULA. If you do not agree to the terms of this EULA, promptly dispose of this copy of the SOFTWARE PRODUCT.

The SOFTWARE PRODUCT is protected by copyright laws and international copyright treaties, as well as other intellectual property laws and treaties. The SOFTWARE PRODUCT is licensed, not sold.

**GRANT OF LICENSE. Prior to Registration.** This EULA grants you the following rights as an unregistered user: For 15 days, you may use the SOFTWARE PRODUCT free of charge. During this time you may make and distribute copies of the SOFTWARE PRODUCT; provided that those copies are unmodified duplicates of the SOFTWARE PRODUCT, and provided that such copying and distribution is made free of charge, and never for profit or for any commercial purpose. The SOFTWARE PRODUCT may not be distributed as part of another commercial product. **After Registration.** This EULA grants you the following rights as a registered user: You, and only you, may use the SOFTWARE PRODUCT on a single computer. This license DOES NOT convey a "site license" allowing others to use the SOFTWARE PRODUCT, or allowing the SOFTWARE PRODUCT to be used on more than one machine simultaneously. You may make and distribute copies of the SOFTWARE PRODUCT; provided (1) that those copies are unmodified duplicates of the SOFTWARE PRODUCT, (2) that those copies DO NOT contain the registration code (normal copying of the SOFTWARE PRODUCT files only will not duplicate the registration code); and (3) that such copying and distribution is made free of charge, and never for profit or for any commercial purpose. The SOFTWARE PRODUCT may not be distributed as part of another commercial product.

**TERMINATION OF LICENSE:** Fifteen days after your first use, your right to use the SOFTWARE PRODUCT under this EULA expires, unless you become a registered user. You may not use the SOFTWARE PRODUCT beyond this expiration, although you may continue to make and distribute copies for others, as described above.

**DESCRIPTION OF OTHER RIGHTS AND LIMITATIONS.** You may not reverse engineer, decompile, or disassemble the SOFTWARE PRODUCT, except and only to the extent that such activity is expressly

permitted by applicable law notwithstanding this limitation. You may not rent or lease the SOFTWARE PRODUCT. Without prejudice to any other rights, Bourbon Street Software may terminate this EULA if you fail to comply with the terms and conditions of this EULA. In such event, you must destroy all copies of the SOFTWARE PRODUCT and all of its component parts.

U.S. GOVERNMENT RESTRICTED RIGHTS. The SOFTWARE PRODUCT and documentation are provided with RESTRICTED RIGHTS. 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 or subparagraphs (c)(1) and (2) of the Commercial Computer Software--Restricted Rights at 48 CFR 52.227-19, as applicable. Manufacturer is Bourbon Street Software, 1568 Englewood Drive, Slidell, LA 70458.

CUSTOMER REMEDIES. The SOFTWARE PRODUCT is supplied "as is." The author and Bourbon Street Software disclaim all warranties, expressed or implied, including without limitation the warranties of merchantability and of fitness for any purpose. The author and Bourbon Street Software assume no liability for damages, direct or consequential, which may result from the use of the SOFTWARE PRODUCT.

# How to contact Bourbon Street Software

We at Bourbon Street Software welcome your comments and suggestions! Internet Marauder was developed with the input of many net surfers looking for a better way, and we look forward to hearing what you have to say about it.

Please feel free to contact us in any of the following ways:

**Email:**      [bourbon@gs.net](mailto:bourbon@gs.net)

**Internet:**    <http://wymple.gs.net/~bourbon>

**Mail:**        Bourbon Street Software  
                 1404 Broadway Street  
                 New Orleans, LA 70118 USA

# Internet Marauder Pro™

If this copy of Internet Marauder isn't enough for you, ask us about our Internet Marauder Pro™ product. Internet Marauder Pro™ sports the following additional features:



Password access -- automatically submits your password for retriected sites!



Multiple retrieval threads per project -- up to 50% faster file retrieval!



Proxy server support -- Use Internet Marauder Pro through firewalls!



File size filters -- retrieve only those files that match a range of sizes that you want!



New Stealth Mode -- aggressively scans html files for links not normally visible to browsers, including commented-out links and links hidden in proprietary scripts!



Two new Marauder personalities -- Global Warlord and Guerilla Fighter!

If you are interested in obtaining Internet Marauder Pro™, please contact [Bourbon Street Software.](#)



**Domain:** the file area contained within a server or a virtual server. The domain of the internet address "www.microsoft.com/home/software/" is "www.microsoft.com." Any internet address beginning with "www.microsoft.com" is located within that domain.

**External domain:** any domain that is not the same as the domain of the starting page.

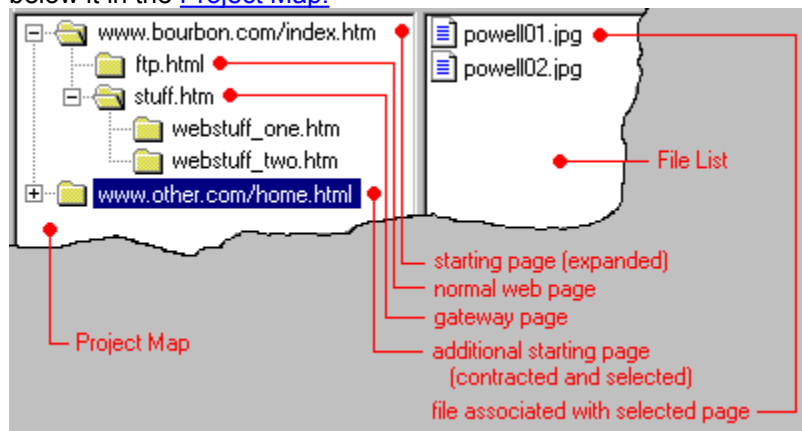
**Embedded graphics** (also **inline graphics**): images (and movies) that appear within a web page, as opposed to those that are linked from a web page (but do not appear on the page when you view it with a browser).

**Toolbar:** a row of buttons for frequently used commands, usually located at the top of the application window. The toolbar is movable and dockable. To move it, grab it on the gray area between and around the buttons and drag it. To dock it, drag it to any window edge until its shape changes to match the side.

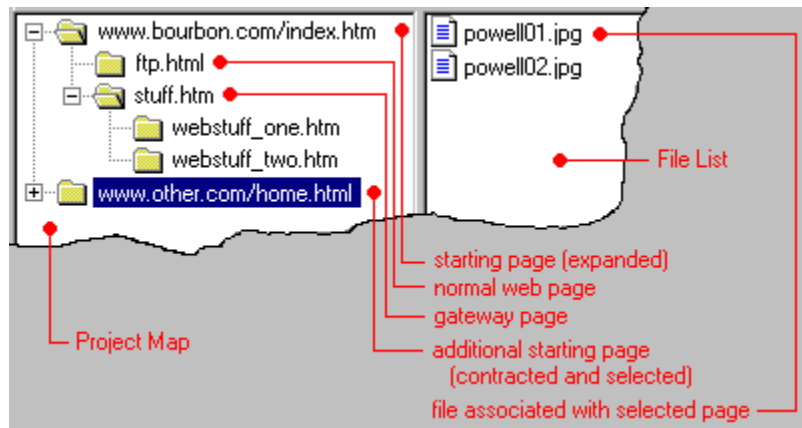


**Spider** (also **robot** or **agent**): a program that travels unattended through a network. Most spiders perform a function at each network location, such as verifying or reporting the location's contents.

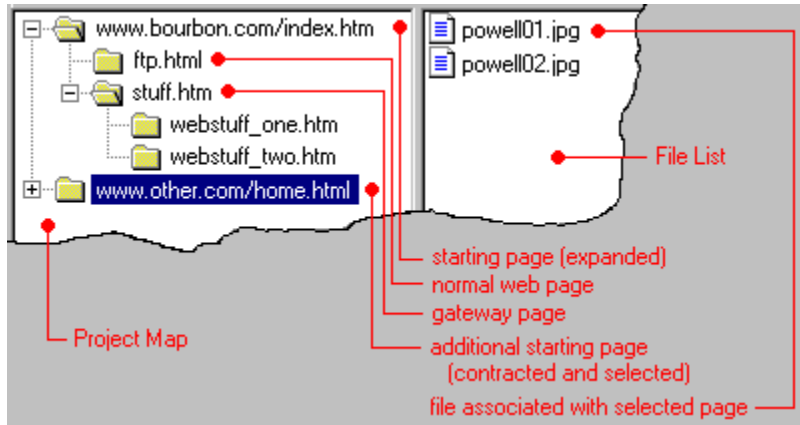
**Starting Page:** The internet address at which the Marauder spider will begin its search. The starting page is entered by the user. It must be the address of a known, operational World Wide Web page. All pages linked from a starting page, either directly or indirectly, "belong" to the page and are indented below it in the [Project Map](#).





**Gateway Page:** The first page that links into an [external domain](#). The gateway page controls Internet Marauder's access into the external domain. By default, whenever Internet Marauder adds a new gateway page to the [Project Map](#), the gateway page is marked "revisit never." This means that Internet Marauder will skip over that domain in later project sessions. To force Internet Marauder to reexplore the domain, change the gateway page's exploration setting.



**Project Map:** the tree view showing web pages that the Marauder spider has visited for the project, displayed in the left-hand pane of the project window. When you select pages in the Project Map, a list of files that were retrieved for that page will appear in the [File List](#). If the web page you select is contracted and it is a [starting page](#) or a [gateway page](#), you will see not only the files that were retrieved for that page, but also the files retrieved for any pages linked from that page.





**File List:** the list view showing retrieved files for the project, displayed in the right-hand pane of the project window. The File List can be switched between a list of names, and a detailed report view, using the   buttons.

