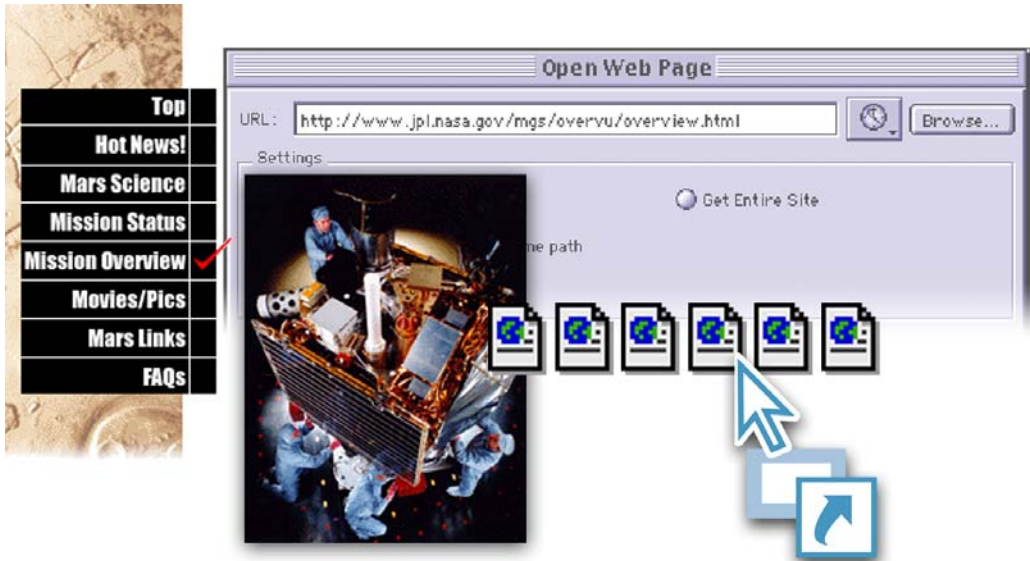


9 Creating Adobe PDF from Web Pages



Acrobat lets you create editable and searchable PDF files by converting Web pages to Adobe PDF files. You can use the resulting PDF files for a variety of archival, presentation, and distribution needs.

In this lesson, you'll learn how to do the following:

- Convert a Web page to Adobe PDF using Web Capture.
- Download and convert Weblinks from a PDF version of a Web page.
- Build a PDF file of favorite Web pages.
- Update or refresh your PDF version of a Web site.

This lesson will take about 35 minutes to complete.

Create a Lesson09 folder on your desktop, in which to save your converted Web pages.

Note: Windows users need to unlock the lesson files before using them. For information, see “Copying the Classroom in a Book files” on page 3.

Converting Web pages to Adobe PDF

You can use Acrobat to download or “capture” pages from the World Wide Web and convert them to Adobe PDF. You can define a page layout, set display options for fonts and other visual elements, and create bookmarks for Web pages that you convert to PDF.


Because captured Web pages are in Adobe PDF, you can easily save, distribute, and print them for shared or future use and review. Acrobat gives you the power to convert remote, minimally formatted files into local, fully formatted PDF documents that you can access at any time.

Web Capture is especially useful for people who make presentations that include Web pages and those who travel a lot. If you need to include a Web site in a presentation, for example, you can convert the Web site to a PDF file so that you have no concern about Web access during your presentation. If you have downloaded all the linked pages, links will behave in the same way as though you were on the actual Web site. Similarly, if you travel extensively, you can create one PDF file containing all of your most visited Web sites. Whenever you have convenient Web access, you can refresh all pages on the site in one simple action. Regardless of the quality or cost of the Web access in your remote location, you can browse your updated PDF versions of the Web sites at your leisure.

Configuring your Internet or proxy settings

Before you use Web Capture, you must configure your Internet or proxy settings for access to the World Wide Web.

- 1 Start Acrobat.
- 2 Choose Edit > Preferences > Internet Settings.
- 3 Do one of the following:
 - In Windows, click the Connections tab in the Internet Properties dialog box, and provide the necessary information for your setup. Your system administrator or ISP will give you the information you need.
 - In Mac OS, select Use an HTTP Proxy Server, and then enter your proxy URL (if necessary) and port number in the text boxes.

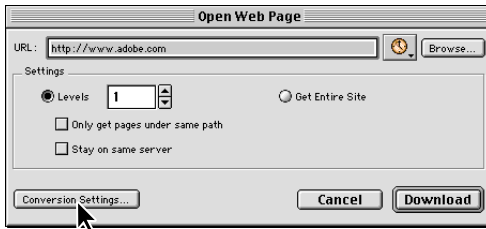
 *In Windows, if you do not configure your Internet settings using the Internet Settings preferences, Internet Explorer must be installed and the Internet Properties dialog box configured to allow access to the World Wide Web. (In an enterprise environment, you are likely to have to configure your proxy server.) Once Internet Explorer has been installed and configured, you may use any browser as your default browser. If your version of Internet Explorer does not have an Internet Properties dialog box, you must upgrade to a current version of Internet Explorer (available from the Microsoft Web site).*

- 4 Click OK to exit the dialog box.

Setting options for converting Web pages

You set options for capturing Web pages before you download the pages. Here, you'll set options for the structure and appearance of your captured pages.

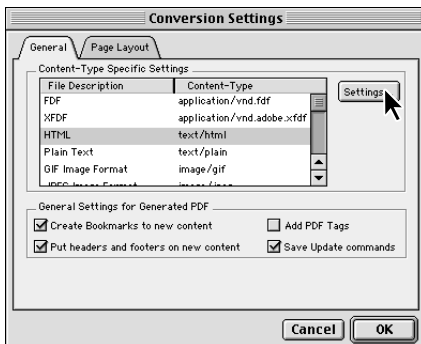
- 1 Choose File > Open Web Page.
- 2 Click Conversion Settings.



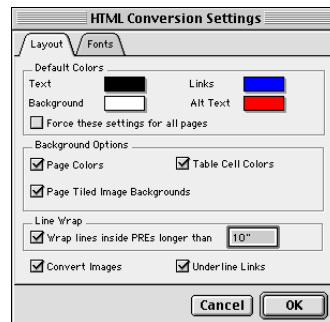
- 3 In the Conversion Settings dialog box, click the General tab.
- 4 Under Content-Type Specific Settings, select HTML and click Settings.

Note: If you have difficulty in Mac OS, choose Help > Adobe Online and check for recent upgrades. Alternatively, reinstall Acrobat 5.0 using the custom installation (rather than the standard installation), and install both Acrobat 5.0 and the Asian language fonts. You will be prompted to restart your computer when the installation is complete.

- 5 Click the Layout tab and look at the options available.



Conversion Settings dialog box



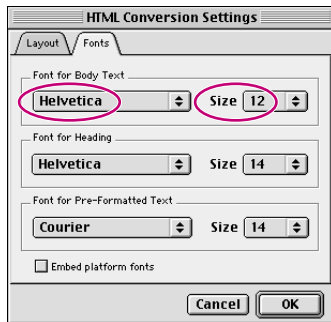
Layout options for HTML conversion

You can select colors for text, page backgrounds, links, and Alt text (the text that replaces an image on a Web page when the image is unavailable). You can also select background display options. For this lesson, you'll leave these options unchanged and proceed to selecting font options.

6 Click the Fonts tab.

7 Under Font for Body Text, do one of the following:

- In Windows, click Choose Font. In the Choose Font dialog box, choose a sans serif font from the Font list. (We chose Helvetica.) Choose 12 from the Size list, and then click OK.
- In Mac OS, choose a font and font size from the pop-up menus. (We chose Helvetica, 12 point.)



8 Under Font for Headings, do one of the following:

- In Windows, click Choose Font. In the Choose Font dialog box, choose a thick sans serif font from the Font list. (We chose Arial Black.) Choose 14 from the Size list, and then click OK.
- In Mac OS, choose a font and font size from the pop-up menus. (We chose Arial Black, 14 point.)

9 Click OK to accept the HTML conversion settings.

10 On the General tab, under General Settings for Generated PDF, select the following options:

- Create Bookmarks to New Content to create a tagged bookmark for each downloaded Web page, using the page's HTML title tag as the bookmark name. Tagged bookmarks help you organize and navigate your captured pages.

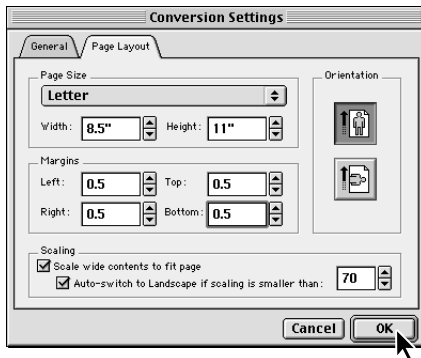
- Add PDF Tags to store a structure in the PDF file that corresponds to the HTML structure of the original Web pages.
- Put Headers and Footers on New Pages (Windows) or Put Headers and Footers on New Content (Mac OS) to place a header with the Web page's title and a footer with the page's URL, page number in the downloaded set, and the date and time of download.
- Save Refresh Commands (Windows) or Save Update Commands (Mac OS) to save a list of all URLs in the PDF file for the purpose of refreshing pages.

11 Click the Page Layout tab.

In Windows, a sample page with the current settings applied appears in the dialog box. You can choose from standard page sizes in the Page Size pop-up menu, or you can define a custom page size. You can also define margins and choose page orientation.

12 Under Margins, enter 0.5 for Left and Right, Top and Bottom.

13 Click OK to accept the settings and return to the Open Web Page dialog box.



After you have some experience converting Web pages to Adobe PDF, you can experiment with the conversion settings to customize the look and feel of your converted Web pages.

Downloading Web pages in Acrobat

You can download Web pages by specifying a URL in Acrobat, by opening a Weblink in a PDF document you already have open, and by dragging and dropping a Weblink or HTML file to an Acrobat window or Acrobat icon. The Web pages are converted to PDF and open in the Acrobat work area.

Note the following when downloading Web pages in Acrobat:

- Acrobat can download HTML pages, JPEG and GIF graphics (including the last frame of animated GIFs), text files, and image maps.
- HTML pages can include tables, links, frames, background colors, text colors, and forms. Cascading stylesheets are supported. HTML links are turned into Weblinks, and HTML forms are turned into PDF forms.

Note: Acrobat downloads the default/index.html frame only once. Other pages may not open in a frame.

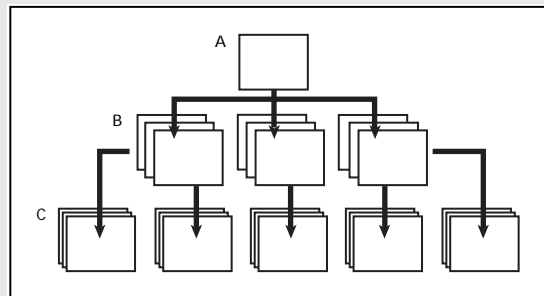
- JavaScript is partially supported at this time; Java applets in HTML pages are not supported.
- To convert Japanese Web pages to PDF on a Roman (Western) system in Windows, you must have chosen to install the Asian language support files at initial installation. (Also, it is preferable to select a Japanese encoding from the HTML conversion settings.) The conversion of Web pages to PDF is not supported for other Asian languages.

About pages on Web sites

Keep in mind that a Web site can have more than one level of pages. The opening page is the top level of the site, and any links on that page go to other pages at a second level. Links on second-level pages go to pages at a third level, and so on. In addition, links may go to external sites (for example, a link at a Web site on tourism may connect to a Web site for a travel agency). Most Web sites can be represented as a tree diagram that becomes broader as you move down the levels.

Important: You need to be aware of the number and complexity of pages you may encounter when downloading more than one level of a Web site at a time. It is possible to select a complex site that will take a very long time to download. Use the Get Entire Site option with great caution. In addition, downloading pages over a modem connection will usually take much longer than downloading them over a high-speed connection.

—From the Acrobat 5.0 online Help.



Web site tree diagram. A First level, B. Second level, C. Third level

Creating an Adobe PDF file from a Web page

Now you'll enter a URL in the Open Web Page dialog box and capture some Web pages.

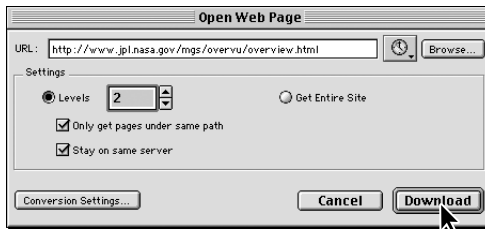
Note: If you are working from within a company network, you may encounter a firewall that limits your access to external Web pages from Acrobat. For instructions on configuring your system to bypass a company firewall, consult your network administrator.

- 1 If the Open Web Page dialog box is not open, choose File > Open Web Page.
- 2 For URL, enter the address of the Web site you'd like to capture. (We used the NASA Mars Global Surveyor site at <http://www.jpl.nasa.gov/mgs/overvu/overview.html>.)

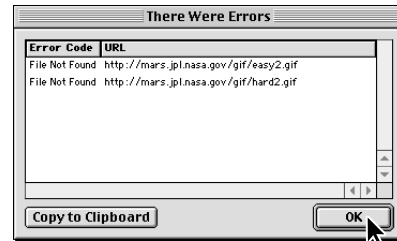
You control the number of captured pages by specifying the levels of site hierarchy you wish to capture, starting from your entered URL. For example, the top level consists of the page corresponding to the specified URL; the second level consists of pages linked from the top-level page, and so on.

- 3 Select Levels, and enter 2 to retrieve two levels of pages in the Web site.
- 4 Select Only Get Pages Under Same Path to capture only pages that are subordinate to the URL you entered.
- 5 Select Stay On Same Server to download only pages on the same server as the URL you entered.
- 6 Click Download. The Download Status dialog box displays the status of the download in progress. When downloading and conversion are complete, the captured Web site appears in the Acrobat document window, with bookmarks in the Bookmarks palette. Tagged bookmark icons differ from the icons for regular bookmarks.

If any linked material is not downloadable you will get an error message. Click OK to close any alert box related to missing material.



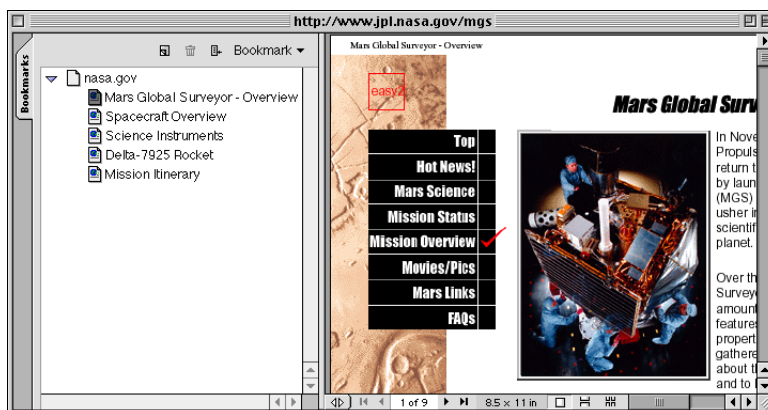
Specifying URL to be downloaded




Error messages indicating missing files

Note: In Windows, if you're downloading more than one level of pages, the Download Status dialog box moves to the background after the first level is downloaded. The globe in the Open Web Page button in the toolbar continues spinning to show that pages are being downloaded. Choose Tools > Web Capture > Bring Status Dialogs to Foreground to see the dialog box again. (In Mac OS, the Download Status dialog box stays in the foreground.)

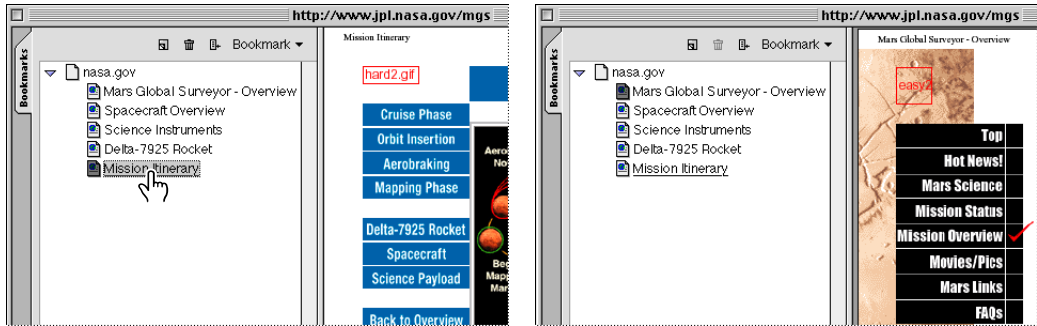
The captured Web site is navigable and editable just like any other PDF document. Acrobat formats the pages to reflect your page-layout conversion settings, as well as the look of the original Web site. Some of the longer Web pages may be spread across multiple PDF pages to preserve the integrity of the page content.



Now you'll use the Bookmarks palette to navigate to another captured page.

7 Select the hand tool () and click a bookmark in the Bookmarks palette.

The page corresponding to the bookmark appears in the document pane.



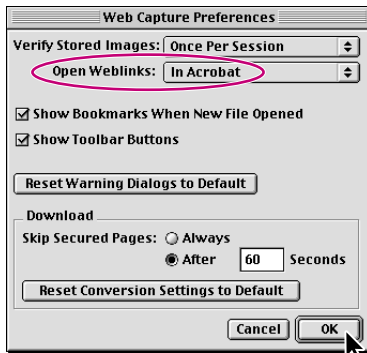
Clicking a tagged bookmark . . .

. . . links to the corresponding page.

Downloading and converting links in a converted Web page

When you click a Weblink in the Adobe PDF version of the Web page that links to an unconverted page, Acrobat downloads and converts that page to Adobe PDF. In order to convert linked pages to Adobe PDF, you must set Web Capture preferences to open Weblinks in Acrobat (the default setting) rather than in your default browser.

- 1 Choose Edit > Preferences > Web Capture.
- 2 For Open Weblinks, choose In Acrobat. Then click OK.



3 Navigate through the captured Web site until you find a Weblink to an unconverted page (we used the link to the Mars Science page), and click the link. (The pointer changes to a pointing finger with a plus sign when positioned over a Weblink.)

Note: If the Specify Weblink Behavior dialog box appears, make sure that Open Weblink in Acrobat is selected, and click OK.

The Download Status dialog box again displays the status of the download. When download and conversion are complete, the linked page appears in the Acrobat window. A bookmark for the page is added to the Bookmarks list.

4 Choose File > Save As, rename the file **Web1.pdf**, and save it in the Lesson09 folder.

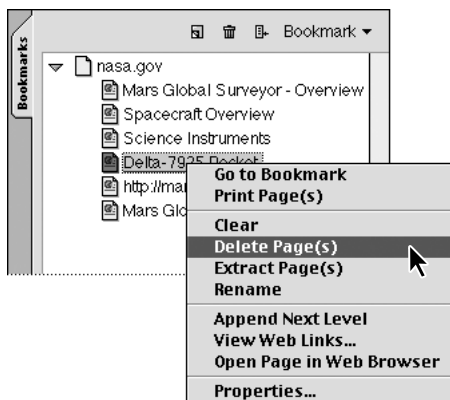
Deleting converted Web pages

Because you downloaded and converted two levels of the Web site in the previous section, you have more converted Web pages than you really want to keep. You'll delete an unwanted page.

1 In the Bookmarks palette, click the bookmark Delta-7925 Rocket to view the related page.

This is the page you'll delete.

2 Right-click (Windows) or Control-click (Mac OS) the bookmark, and choose Delete Page(s) from the context menu. Click OK in the alert box.



Note: If you click Delete (Windows) or Clear (Mac OS) in this context menu, you delete the bookmark only.

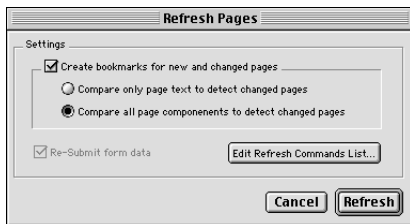
Updating converted Web pages

Because you selected the Save Refresh Commands option when you first converted the Web pages to Adobe PDF, you can refresh or update all pages from one or multiple sites from one Acrobat dialog box.

You can refresh Web pages in a PDF document to retrieve the most up-to-date version from the Web site. Whenever you use the Refresh command, you download the entire Web site or link again and build a new PDF file. Any pages where components have changed—for example, text, Weblinks, embedded filenames, and formatting—are listed as bookmarks in the Bookmarks pane under the New and Changed Pages bookmark. Any new pages that have been added to the site are also downloaded.

Note: The Refresh command may not update converted Web pages that contain forms data. In this case, you will get an error message identifying the pages.

- 1 With an internet connection open, choose Tools > Web Capture > Refresh Pages.
- 2 Select Create Bookmarks for New and Changed Pages.



You specify whether Acrobat looks only for text changes or all changes, including text, images, Weblinks, embedded files, etc. (We used Compare All Page Components to Detect Changed Pages.)

- 3 Click Edit Refresh Commands List.

This window displays the URLs of all the Web sites that have been converted to Adobe PDF in this file. You can deselect any URLs for pages that you don't want to refresh.



Select the URLs for pages you want to refresh.

4 Click OK to accept the default selection, and click Refresh to update your converted PDF pages.

5 Choose File > Save. Notice that this is a new file. If you want to keep the Web1.pdf file for archival purposes, save this file under a different name.

Earlier in the lesson, you deleted a page from the PDF file of the converted Web site. Notice that the page you deleted in the earlier version is present in the refreshed file. Any reorganization or deletion of pages in the PDF file is lost when you refresh.

Building an Adobe PDF file of favorite Web pages

Earlier in the lesson you created a PDF file containing Web pages on the Mars Global Surveyor and its mission. In addition to surfing this site, suppose you also like to check the Adobe home page.

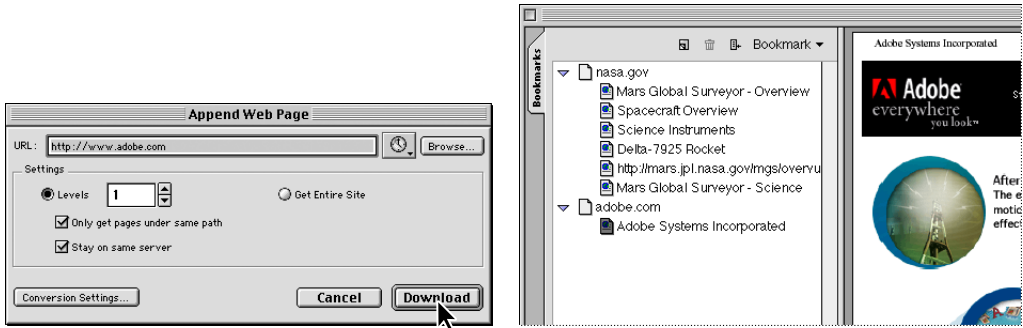
In this section, you'll start to build an Adobe PDF file of your favorite Web pages by appending the Adobe home page to the Mars Global Surveyor file that you have already created.

1 With the Web1.pdf file open, choose Tools > Web Capture > Append Web Page.

2 In the Append Web Page dialog box, enter the URL for your favorite Web site. (We used www.adobe.com.)

3 For Levels, select 1. Select the Only Get Pages Under Same Path option and the Stay On Same Server option.

4 Click Download.



Acrobat appends a PDF version of the Adobe home page to the Mars Global Surveyor pages. You can download any linked pages on this site as described in “Downloading and converting links in a converted Web page” on page 248.

You can repeat this process to build up a file of your favorite Web sites that you can keep on your system, share with friends, and refresh at your convenience.

Exploring on your own (Windows)

Earlier in the lesson you created a PDF file that contains several converted Web pages related to the Mars Global Surveyor and its mission. You can easily and quickly add unrelated pages to this file by dragging and dropping links into the PDF file.

To start building your customized PDF file of Web pages, you’ll add a page from the Adobe Web site.

- 1 If needed, open the Web1.pdf in the Lesson09 folder.

2 In your browser, go to the Adobe home page (www.adobe.com). Resize the browser window so that the Acrobat window is visible beneath.



3 Browse the Adobe home page to find a link. (We used the Free e-mail newsletters link under Adobe announces.)

4 Drag the link from the Adobe home page onto the Acrobat window. Release the mouse button when the cursor icon changes to a link icon.

The Web page represented by the link is converted to PDF and appended to the end of your Mars Global Surveyor file.

5 Close the browser window.

6 Choose File > Save to save the file if you wish.

Review questions

- 1 How do you control the number of Web pages captured by Acrobat?
- 2 How do you convert destinations of Weblinks to PDF automatically?
- 3 How do you update your PDF file to show the latest version of the captured Web site?

Review answers

1 You can control the number of captured Web pages by specifying the following options:

- The Levels option lets you specify how many levels in the site hierarchy you want to capture.
- The Only Get Pages Under Same Path option lets you download only pages that are subordinate to the specified URL.
- The Stay On Same Server option lets you download only pages that are stored on the same server as the specified URL.

2 To convert the destination of a Weblink to PDF, first choose Edit > Preferences > Web Capture, and choose In Acrobat for the Open Weblinks option. Then click the Weblink in the PDF file to convert the link's destination to PDF.

3 With an internet connection open, choose Tools > Web Capture > Refresh Pages to build a new PDF file using the same URLs and links. Select the Create Bookmarks for New and Changed Pages option if you want Acrobat to create bookmarks for pages that have been modified or added to the Web site since you last converted the Web site and its links. You also specify whether Acrobat looks only for text changes or for all changes. (You must have selected the Save Refresh Commands (Windows) or Save Update Commands (Mac OS) option in the Conversion Settings dialog box to save a list of all URLs in the PDF file for the purpose of refreshing pages.)