Scanning guide for

ServiceLogic

Pre-release document
January 2004

Contents

Storing paper documents in ServiceLogic

ServiceLogic provides a document repository for your electronic documents. However, many critical documents exist as paper documents only, and the electronic document is not available. To store a paper document in ServiceLogic, you need to scan it with a document scanner and save it in an electronic format.

There are too many different scanners and scanning software programs to discuss them all in this guide. So instead, this guide is presenting a discussion of what factors to consider when scanning documents, as well as some descriptions of our experiences with several specific tools that worked well for us.

In general, to scan electronic documents you need a scanner, software to control the scanner and create the electronic files. Our preferred setup includes the following:

- Scanner - Fujitsu fi-4120c. A relatively high-speed scanner with a document feeder.
- Software - ABBYY FineReader. A program that accepts images from the scanner and converts them to searchable PDF files.

Searchable documents vs. image-only documents

Electronic documents that are created by programs such as Word or Excel are searchable documents. "Searchable" means that certain software programs can read the text of the document and use it to find the document. This is how you can use ServiceLogic's Search feature to find a document based on keywords within the document.

But when you scan a paper document, the document is stored as an image-only document (unless you process it with additional software).

Unlike a spreadsheet or a word-processing document, an image-only document just contains a picture of the page. Like a fax, an image-only document does not contain the actual data that describes those words on the page. Therefore, software programs (including ServiceLogic's Search feature) cannot search keywords within an image-only document.

Converting images to searchable documents (OCR)

To convert an image-only document to a searchable document, you need special software that performs Optical Character Recognition (OCR). OCR software "reads" the image by looking at the shape of each word and letter.

This guide covers two ways to create searchable documents for ServiceLogic

- **Searchable PDF files.** You can use special software on your computer to perform OCR on the scanned image, and save the document in a searchable PDF format. You can then use the ServiceLogic Search feature to search for keywords within searchable PDF files.
- **.tif files**. If you save your scanned document as a .tif file, ServiceLogic performs the OCR for you and saves the resulting keywords in a special index. As a result, you can use the ServiceLogic Search feature to search for keywords within the document. There are advantages and disadvantages to using .tif files - for more information, see TIF Files later in this guide.

PDF files

A PDF file (Portable Document Format) is a special file format that is specified by Adobe Systems, Inc. (www.adobe.com). PDF is widely used as a common format for sharing documents.

The wide availability and familiarity with PDF files makes it a good choice for documents that need to be printed and shared. Any document can be converted to PDF format, and the software to read PDF files is free from Adobe and a number of other providers. In fact, your browser is probably already configured to open and display PDF files.

Searchable PDF files

If you scan a document without performing OCR, you can save the scanned image as a PDF file. However, the resulting PDF file is an image-only format, and you cannot search for keywords within the text. But if you use special software such as Acrobat Capture, or ABBYY FineReader, you can perform OCR on the image and create a type of file called "searchable PDF."

A searchable PDF file is a very clever way of solving the problem of searching for text within a scanned image, yet still retaining the ability to print an accurate copy of the scanned page. To create a searchable PDF file from a scanned image, your software stores the OCR'd text in a hidden index in the PDF file. When you print a searchable PDF or view it on the screen, you are looking at the scanned image of the page, just like a copy or a fax. But when you search for keywords within the text, the search looks at the hidden index of text.

Most scanners are bundled with some sort of OCR software. However, for our purposes we need an OCR program that can also create searchable PDF files. The software that comes with most scanners cannot create searchable PDF files.

ABBYY FineReader (OCR software with searchable PDF creation)

For the purposes of this guide, we informally evaluated several OCR tools capable of creating searchable PDF files. We chose ABBYY FineReader because it is excellent software, it appears to have the lowest cost, and was the easiest to use. The version we evaluated was: "ABBYY FineReader 7.0 Professional Edition."

ABBYY FineReader is available from ABBYY Software House (www.abbyy.com). You can buy a full version online, or you can buy a copy from several resellers. A 30-day trial download is available.

The cost is $299.99 USD for the full version. However, an "Upgrade Version" is available for $149.99, and the upgrade terms are very lenient. Here is the exact quote from their web site:

> "UPGRADE price applies to existing user of all OCR software. If you own any OCR software, you qualify for this price!"

Most likely, the OCR software that came with your scanner makes you eligible for the upgrade price.

ABBYY FineReader is available for Windows 98/2000/XP/NT4 and Macintosh. It supports most TWAIN-compliant scanners (most scanners are TWAIN-compliant). More information is available on the ABBYY website, including a list of supported scanners and system requirements.

About Adobe Acrobat

Acrobat 6.0 has similar functionality via its Paper Capture feature, although Acrobat is more expensive and seemed slightly more difficult to use. If you already own Acrobat 6.0, or if it comes with your scanner, you may want to consider using Acrobat instead of ABBYY FineReader.

Installing ABBYY FineReader

For our evaluation, we downloaded the full-featured trial version of ABBYY FineReader from the ABBYY website at http://www.abbyy.com.

ABBYY FineReader has a typical Windows installation program that is very easy to run. It detected our scanner automatically, and no further configuration was necessary.

ABBYY FineReader supports most currently available scanners. However, if ABBYY cannot detect your scanner, check the ABBYY web site to see if your specific scanner is supported.

How good is OCR?

OCR programs have been available for many years, and in recent years they have become much better and more accurate (and also less expensive). OCR is not yet perfect, although it can reliably reach accuracy rates of 98 - 99 percent or higher.

Fortunately, that accuracy rate is more than good enough for our purposes. That's because when we create a searchable PDF (or when ServiceLogic creates an OCR'd index for a .tif file), the OCR'd text is always hidden. We don't need to look at it at all - it's just there so we can find the keywords. But when we display or print the document, what we are printing is the image, not the OCR'd text.

We aren't trying to use the OCR'd text to recreate the formatted document. As long as the OCR process gets most of the words right, that is adequate for creating a searchable index.

Most OCR software, including ABBYY FineReader, provides an opportunity for you to preview and correct any errors the OCR process has made before you save it. In addition, most OCR software allows you to set various configuration options to make the OCR more accurate. You can do this if you find it necessary, but these additional settings are not covered in this guide.

If you don't use OCR software

If you don't have OCR software, or just want to avoid the additional effort of creating searchable documents, you may be satisfied with image-only documents.

- Image-only documents are perfectly adequate for archival purposes, since you can store them in ServiceLogic and print them out again.
- In ServiceLogic you can still search for image-only documents based on the document name and on document notes (you can enter document notes when you upload the document).
- If you create .tif files, ServiceLogic will perform the OCR for you, so the document will be searchable by keyword (but only within ServiceLogic).

Choosing a scanner

The good news is that just about any currently available scanner will produce images that are more than acceptable for scanning business documents. You should select a scanner based on other criteria, such as:

- Scanning speed
- Document feeder
- Available support

Many scanners, especially scanners that are sold in office superstores, are aimed at home users who are primarily interested in scanning photographs. The consumer scanners usually produce perfectly acceptable images, but they tend to be slower, they tend not to have document feeders, and they tend to be bundled with "lite" OCR or other consumer-oriented programs such as photo albums that are not needed to operate the scanner.

Flatbed vs. sheetfed scanners

A flatbed scanner has the advantage that you can scan bulky items like books, binders, or stapled pages. However, you have to put each page on the scanner glass by hand.

Some scanners are sheetfed, which means that you can feed one or more sheets of paper through the scanner's document feeder. This is a significant advantage if you are scanning a lot of standard-sized loose sheets of paper.

Some sheetfed scanners have both features - they contain a document feeder mounted on top of a flatbed scanner.

The Fujitsu fi-4120c scanner described in this guide is a sheetfed-only scanner, which means that we cannot scan books or bulky items on it.

One option might be to buy a fast sheetfed scanner for most of your documents, and to also buy an inexpensive flatbed scanner for occasional bulky items.

Duplexing scanners

Some scanners are duplexing scanners, which means the scanner scans both sides of a sheet of paper at the same time. If you scan a lot of two-sided paper, a duplexing scanner can be a major convenience feature.

Scanning speed

Scanning speed is an important consideration. Even the most inexpensive scanners can produce good images, but many of them scan at a fairly slow rate. A slow speed may be acceptable to you if you don't do much scanning, but if you scan every day, or if you need to scan many pages, we suggest you find a scanner that scans at a higher speed.

Scanning speed is a factor that can significantly increase the price of the scanner. However, if you do a lot of scanning you may decide it is worth it.

Support

You may wish to find a local office machine reseller who will provide ongoing support to install the scanner and troubleshoot if necessary.

Dots per inch

Your scanner must be able to scan at a resolution of at least 300dpi (dots per inch). Even the least expensive scanners on the market are able to scan at least at 600dpi, and some much higher.

Your scanner's dpi capability determines how much detail the scanner is able to capture from the page. 300dpi is the optimum setting for performing OCR - more resolution is not necessary and may even make the OCR process more difficult. When you perform the scan, you can configure it to scan at 300dpi, no matter how many dpi your scanner is capable of.

Faxes are typically scanned at 200dpi (although some are now higher). 200dpi produces an acceptable printed image for most business documents, but 300dpi is recommended for OCR.

USB connection

USB (Universal Serial Bus) is a type of connection found on most PCs. Most scanners can connect to your PC's USB port.

Fujitsu fi-4120c scanner

This guide were prepared using a Fujitsu 4120-C. The device was ordered from A Matter of Fax of 105 Harrison Avenue, Harrison, NJ 07029 (phone 973-482-3700). The cost was $887.87 with shipping in October, 2003.

The advantages of this device are a small footprint, fast scanning with a 50 sheet capacity feeder, duplex capability and support for both TWAIN and ISIS drivers. The machine comes with:

- a fast SCSI adapter for use in desktop machines
- Adobe Acrobat 5.0
- SCSI connector cable
- USB connector cable
- User manual and a “quick setup guide” chart
- Driver software on a CD

Figure 1. Fujitsu fi-4120c document scanner

Storing electronic documents in ServiceLogic

Any electronic document can be uploaded to ServiceLogic.

ServiceLogic's search feature can find documents based on three different types of criteria:

- Document name - the name you gave the document when you uploaded it.
- Document notes - any comments you entered when you uploaded the document.
- Keywords - keywords within the document.

The search feature is able to read keywords within most popular document formats, including Microsoft Office documents such as Word and Excel.

ServiceLogic can read keywords in the documents because you (or somebody else) typed the words into the document, and the document contains all the words. However, if the document is scanned and saved as an image, ServiceLogic cannot search keywords within the image unless you perform some further processing to convert the image to a searchable format. There are several ways to perform this conversion...

About .tif files

Probably the simplest way to save scanned documents is to save them as .tif files. .tif files are an image-only format. However, when you upload a .tif file, ServiceLogic automatically performs OCR on the image and creates a searchable index of the document text. Once ServiceLogic OCRs and indexes the image, you can search for keywords within the text to find the uploaded .tif file. ServiceLogic does not perform OCR for any format other than .tif.

Advantages

To create .tif files, you don't need OCR software (such as ABBY FineReader). You can create .tif files with commonly available scanning software. Most likely, your scanner was bundled with some basic software that can create .tif files. Windows 2000 also includes a program called "Imaging for Windows" that can create .tif files. There are also free scanning programs available from shareware sites.

Disadvantages

- .tif files can have a somewhat larger file size than .pdf files.
- .tif files are unfamiliar to many people. If you plan to send the electronic .tif files to other people, they may not recognize it as a standard document format, and their systems may not be configured to open the file. However, if you just want the document archived quickly and don't plan to distribute it further, a .tif file is probably the fastest and simplest type to create.

About the .tif format

.tif and .tiff, TIF and TIFF are all the same image format, and most Windows programs will accept any version of the file extension. TIFF, which stands for Tagged Image File Format, was originally created for Unix systems, where four-character file extensions are permitted. However, when the TIFF format began to be used on DOS and early versions of Windows, only three letters were permitted. Therefore it was necessary to shorten the extension to three letters for DOS and Windows programs. Current versions of Windows do not have this restriction, so you can use any version of the extension (.tif and .tiff, TIF and TIFF).

.tif file size

Within the .tif file specification, there are several different sub-formats. The details are beyond the scope of this guide, but our recommendation is that .tif files should be saved as "Black and white, Type 4." This format will create the smallest file size. Other .tif formats can be uploaded and OCR'd by ServiceLogic, but the other .tif formats will unnecessarily take up extra file space.

Scanning a .tif file with Imaging for Windows

Imaging for Windows (IFW) is a basic scanning program included with Windows 2000. It is apparently not included with Windows XP. However, other basic scanning programs should follow a similar workflow, so you can use the following procedure as a guideline.

Before beginning, make sure you turn on the power to your scanner.

Open Imaging for Windows

1. On the Start Menu, click Programs | Accessories | Imaging. Imaging for Windows (IFW) opens.

Create a new empty .tif document

IFW requires you to create a new empty document before scanning, in order to set the correct configuration.

2. On the File menu, click New. The New Blank Document dialog opens.
3. In the New Blank Document dialog, select the following options:
   - On the File Type tab, select TIFF Document.
   - On the Color tab, select Black and White.
   - On the Compression tab, select "CCITT Group 4 (2d) Fax" This format produces the smallest .tif files. Other formats will work, but will result in a larger file size.
   - On the Resolution tab, select "300x300 dpi" if you think you might want to OCR the image in the future. This resolution is optimal for OCR. At lower resolutions you may be satisfied with the print quality of the image, but OCR may not be reliable below 300 dpi.
4. Click OK. The New Blank Document dialog closes.
5. On the File menu, click Save. Make sure you save the document as a .tif file.

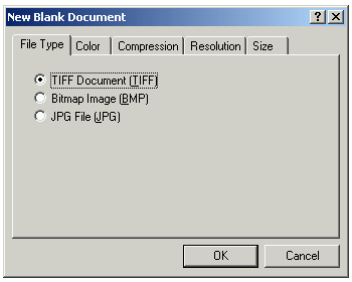

Figure 2. New Blank Document dialog

Open your scanner's TWAIN Driver dialog

6. On the File menu, click Acquire Image. Your scanner's TWAIN dialog opens.

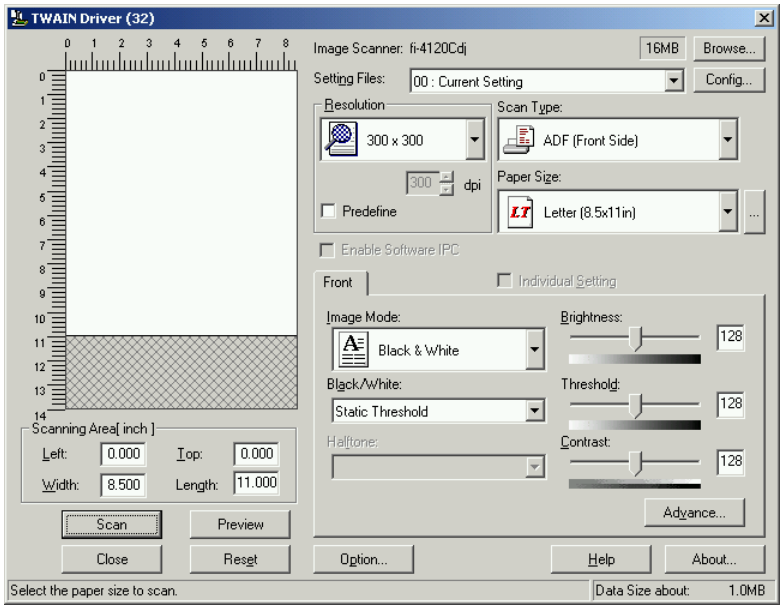

Figure 3. TWAIN dialog for the Fujitsu 4120c

Review your scanner settings

7. In the scanner's TWAIN Driver dialog, review all scanner settings.

   - Resolution - 300 dpi (300x300). This is the optimum resolution for OCR. Higher resolutions are not recommended.
   - Image mode - Black & White. Most business documents should be scanned in black and white to create the smallest file size. You can scan in color if you like, but that is not in the scope of this guide.
   - If your paper documents are two-sided, choose ADF (Duplex) in the Scan Type list. The Fujitsu 4120c is a duplexing scanner, which means it can scan both sides of a page on the same pass through the machine. If your scanner is not a duplexing scanner, this option is not available.
   - For more information about other scanner settings, click the Help button.

Load your paper documents in the scanner

8. Make sure the paper documents are loaded in the scanner.

Start the scan

9. In the TWAIN Driver dialog, click Scan. The TWAIN Driver dialog closes, and your scanner should start scanning the documents.

Review the scanned image

When the scanner is finished, the scanned image appears in the main window of IFW.

10. If some pages are upside down, you can rotate them to turn them right-side up.:

- If all the scanned pages are upside down, on the Page menu click Rotate All Pages | 180°
- If only some of the scanned pages are upside down, review each page one at a time, and on the Page menu click Rotate | 180° for each page.

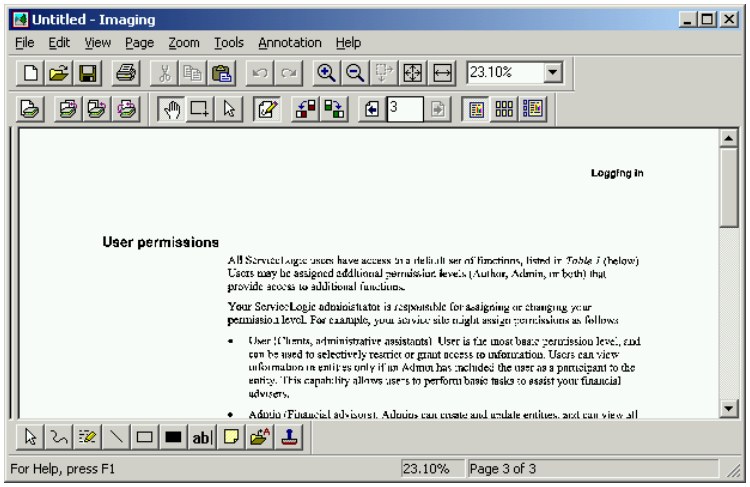

Figure 4. Imaging for Windows

Save the scanned image as a .tif file

11. On the File Menu, click Save. The Save As dialog opens. Make sure the file type is .tif, then save the file.

After you save the file, the file is ready to be uploaded to ServiceLogic.

Scanning a .PDF file with ABBYY FineReader

Before beginning, make sure you turn on the power to your scanner.

Open ABBYY FineReader

1. On the Start Menu, navigate to ABBYY FineReader and start the program.

Open your scanner's TWAIN Driver dialog

2. On the Process menu, click Scan & Read Multiple Images if you are scanning more than one page. Your scanner's TWAIN Driver dialog opens.

- If you are scanning only one page, click Scan & Read instead. The procedure is similar, except that the TWAIN Driver dialog closes automatically after scanning, but for Multiple Images the dialog remains open until you close it.

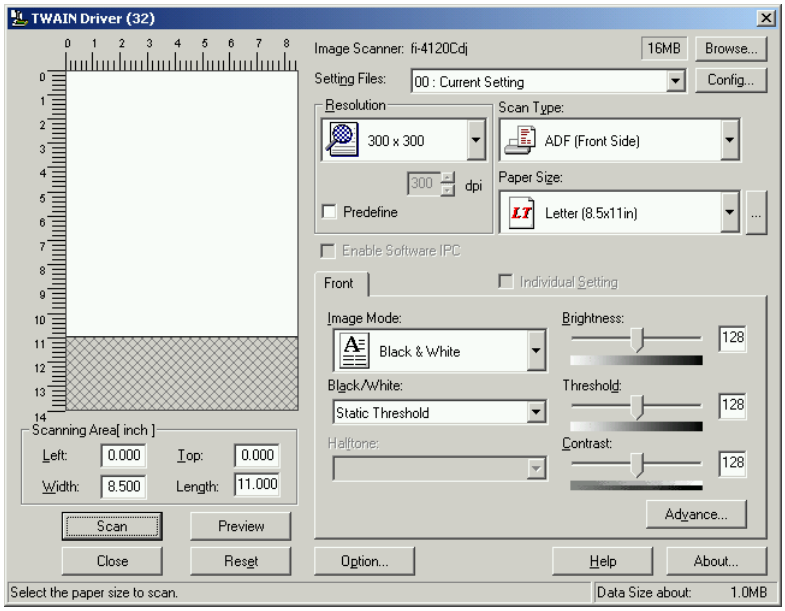

Figure 5. TWAIN dialog for Fujitsu 4120c

Review your scanner settings

3. In the scanner's TWAIN Driver dialog, review all scanner settings.

   - Resolution - 300 dpi (300x300). This is the optimum resolution for OCR. Higher resolutions are not recommended.
   - Image mode - Black & White. Most business documents should be scanned in black and white to create the smallest file size. You can scan in color if you like, but that is not in the scope of this guide.
   - If your paper documents are two-sided, choose ADF (Duplex) in the Scan Type list. The Fujitsu 4120c is a duplexing scanner, which means it can scan both sides of a page on the same pass through the machine. If your scanner is not a duplexing scanner, this option is not available.
   - For more information about other scanner settings, click the Help button.

Load your paper documents in the scanner

4. Make sure the paper documents are loaded in the scanner.

Start the scan

5. In the TWAIN Driver dialog, click Scan. The scanner should start scanning the paper documents. FineReader begins OCR processing of each page shortly after it is scanned.

   - If the page is upside down, FineReader rotates it to the correct orientation. You do not need to rotate the pages yourself.

6. After all the documents have passed through the scanner, click Close to close the TWAIN Driver dialog.
    - You can feed more documents into the scanner if necessary. All the scanned documents will be added to the same batch (and added to the same PDF file) until you close the TWAIN Driver dialog.

Review the scanned image

7. When the scanning is finished, the scanned image appears in the main window of ABBYY FineReader.

Save the scanned image as a .PDF file

8. On the File Menu, click Save Text As. The Save Text As dialog opens. Make sure the following options are selected, then click Save:
    - Save as type: PDF Document
    - Save pages: All pages
    - File options: Create a single file for all pages

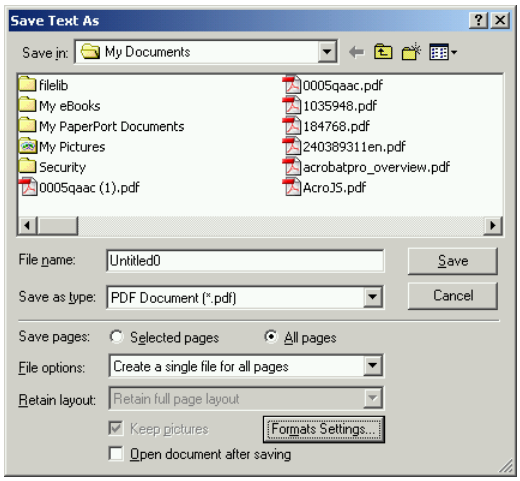

Figure 6. Save Text As dialog for ABBYY FineReader

9. In the Save Wizard, click the Formats Settings button. The Formats Settings dialog opens. On the PDF tab, make sure the Save Mode is "Text under the page image."
    - This setting produces the searchable PDF file that displays and prints the image, while retaining the hidden OCR'd text index.
    - This setting appears to be "sticky," so after you set it once this option is retained and you do not need to reset it every time you scan a document.:

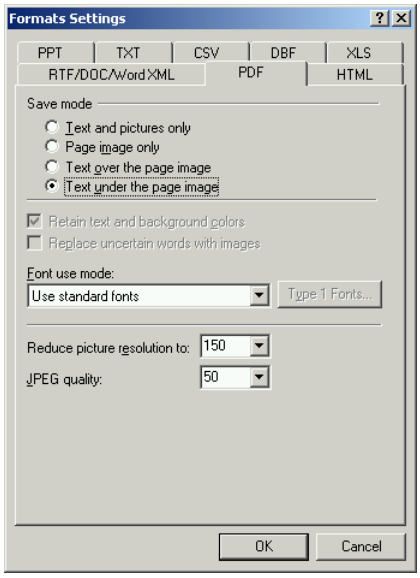

Figure 7. Formats Settings dialog for ABBYY FineReader

10. Click OK to close the Formats Settings dialog, then click Save (on the Save Text As dialog) to save the file.

After you save the file, the file is ready to be uploaded to ServiceLogic.

Installing the scanner and OCR software

These instructions are based on our experiences installing the Fujitsu fi-4120c and ABBYY FineReader on Windows 2000. The installation sequence may be different if you have another version of Windows, or some other differences in your environment.

Here are some highlights of lessons learned while setting up our scanner:

- No matter which scanner you are using, you should check with the vendor to make sure you have the most recent driver software. We initially installed the driver from the CD-ROM that came with the Fujitsu scanner, and we immediately experienced a number of minor problems. But when we downloaded the latest driver from Fujitsu, all the problems were resolved.
- Think twice before installing all the software that came with your scanner. You probably don't need most of it. Many scanners are bundled with "lite" OCR or other consumer-oriented programs such as photo albums that are not needed to operate the scanner.
- We ended up not using any of the software packaged with the Fujitsu scanner. We downloaded the latest driver from Fujitsu, and we downloaded ABBYY FineReader from the ABBYY web site.
- To operate your scanner, all you need is the scanner driver, and some scanning software (for example, ABBYY FineReader, or Adobe Acrobat).
- When we installed ABBYY FineReader, it found the Fujitsu scanner automatically, and no further configuration was needed.

Summary of installation process

Here is the installation process in summary:

1. If the scanner is already installed, uninstall it.
2. Make sure you have the latest TWAIN drivers for the scanner and save them in a folder. Uncompress the driver file to a folder, and run the driver's setup program to automatically copy the drivers to your Windows system directory.
3. Connect the scanner to the USB port and turn on the power. Windows should detect the presence of the scanner and install it automatically, using the drivers from the Windows system directory. If it doesn't detect the scanner, try rebooting the computer.

Installing the scanner

If you have installed the driver software as described in the topic "Getting the latest driver for your scanner," then Windows selects the correct driver automatically the first time you connect the scanner, or when you restart your computer.

Connect your scanner

1. Plug in your scanner's power cord and make sure the power is turned on.
2. Connect the scanner's USB cable to a USB port on your PC.
    - The Fujitsu scanner also supports a SCSI connection, but we used USB for our installation. USB is more common and is easier to configure.
3. Windows detects your scanner and installs the correct drivers. If your scanner is not detected, try restarting your computer.

Getting the latest driver for your scanner

No matter which scanner you are using, you should check with the vendor to make sure you have the most recent driver software. We initially installed the driver from the CD-ROM that came with the Fujitsu scanner, and we immediately experienced a number of minor problems. But when we downloaded the latest driver from Fujitsu, all the problems were resolved.

At the time of this writing, the latest driver for the Fujitsu 4120c is available from the driver download page at:

http://www.fujitsu.com/support/computing/peripheral/scanners/drivers/twain91147

You'll need to choose the driver for your operating system. For this demonstration, we are using Windows 2000.

The driver download page contains detailed instructions for installing the scanner. Here is a summary based on our experiences:

1. Save the downloaded file on your PC. The downloaded driver file is ft9b1bxb.exe. This is a self-extracting compressed file.
2. Copy or move the downloaded file to a new folder. Call the folder something like "fi_drivers."
3. Double-click the file. The file uncompresses itself to create a folder named Disk 1.

4. In the Disk 1 folder, click setup.exe. A typical Windows installation program opens. This installation program installs tools and drivers that your scanner needs. In particular, it copies the basic driver files to your Windows system directories. When you connect your scanner (or perhaps when you restart your computer), the Windows plug and play process should detect the scanner and automatically install the correct drivers.

Uninstalling your scanner

1. On the Start menu, click Settings, then click Control Panel. The Control Panel opens.
2. In the Control panel, double-click Scanners and Cameras. The Scanners and Cameras Properties dialog opens.
   - If you don't see the Scanners and Cameras listed in the Control Panel, you probably don't have any scanners or cameras installed.
3. Click the name of the scanner, then click Remove

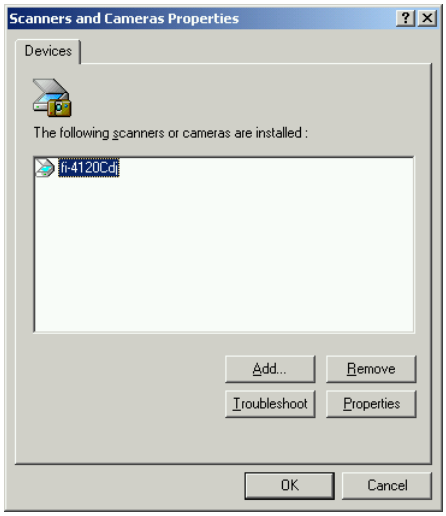

Figure 8. Scanners and Properties dialog for the Fujitsu 4120c