

Building a Searchable PDF Library and Catalog



Converting all of your electronic and paper publications to PDF lets you distribute and search large collections of documents quickly and easily. You can use Acrobat Catalog to create a full-text index of your PDF publications, and then use the Search command in Acrobat or Acrobat Reader to search the entire library almost instantly.

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

- Build an index using Acrobat Catalog.
- Use the Search command to locate information contained in the files indexed by Catalog.
- Set Search options.
- Refine a search.
- Use Document Info fields to conduct a search.
- Search for information using Boolean expressions.
- Choose Catalog options.
- Search for a phrase.

This lesson will take about an hour to complete.

If needed, remove the previous lesson folder from your hard drive, and copy the Lesson11 folder onto it.

Building an index

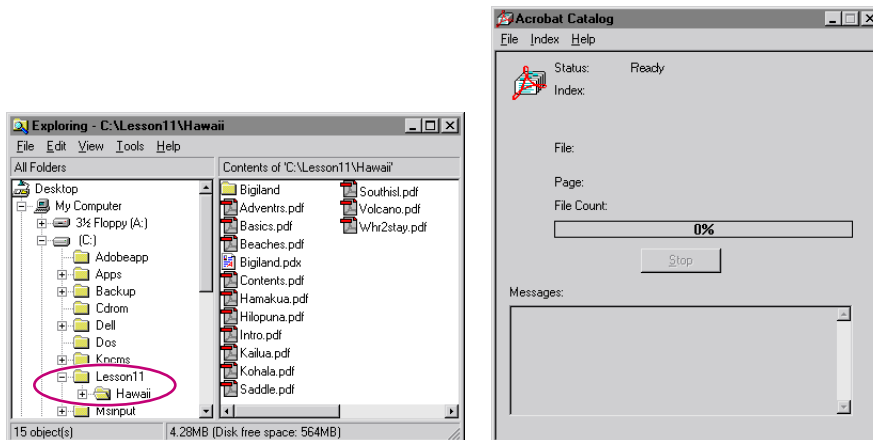
You use Acrobat Catalog to build full-text indexes of PDF document collections. A full-text index is a searchable database of all the text in a document or set of documents. Your documents should be complete in content and electronic features such as links, bookmarks, and form fields before you use Catalog to index them. In this lesson, you'll work with chapter files from the book, *Hawaii: The Big Island Revealed*, by Andrew Doughty and Harriett Friedman.

You'll create an index of these files and then search that index to find exciting information about what to do on your next trip to Hawaii.

1 From the desktop, open the Hawaii folder in the Lesson11 folder, located inside the Lessons folder within the AA4_CIB folder on your hard drive. Notice the files contained within this folder. All of the PDF files in this folder will be indexed by Catalog.

Before you index a document collection, you need to organize the documents on the disk drive or network server volume, make sure the filenames comply with cross-platform conventions, break large documents up into smaller files (to enhance search performance), and complete Document Info fields in each document, if appropriate. (See “Using Document Info fields to search” on page 269.)

2 Start Acrobat Catalog.



Files contained in Hawaii folder

Acrobat Catalog

3 In Mac OS, choose Edit > Preferences, make sure that Make Include/Exclude Folders DOS Compatible is deselected, and click OK.

4 Choose Index > New (Windows) or File > New (Mac OS). For Index Title, enter **Hawaii: The Big Island**.

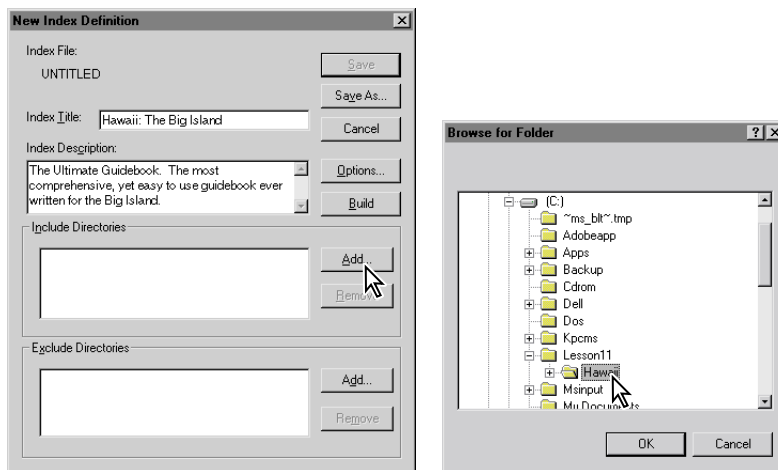
The Index Description box provides users with more information about the documents included in the index.

5 Click inside the Index Description box, and enter information about the index you are building. (We entered “The Ultimate Guidebook. The most comprehensive, yet easy to use guidebook ever written for the Big Island.”)

6 Under Include Directories, click Add.

Now you'll select the folder or folders that contain the documents to be indexed.

7 Click to select the Hawaii folder in the Lesson11 folder, and then click OK (Windows) or click Select "Hawaii" (Mac OS).



Entering index description

Folder where you'll save results

By default, Catalog indexes subdirectories, but you can exclude subdirectories using the Exclude Directories option. You can also add more than one directory to the index. For this lesson you'll index one directory.

8 Click Build.

9 Open the Hawaii folder so that you can save the index within that folder. Name the index **Guide.pdx**, and click Save. The file extension PDX identifies a file as an Acrobat index.

To make searches more efficient, you can save the index within the folder of information it's indexing. In Windows, the folder must be on the disk or network server volume where the documents to be indexed are stored. In Mac OS, the folder may be put on a different disk or network server volume from that of the indexed documents, if you don't plan to move the index and documents. (In that case, you would choose Edit > Preferences > Index > Allow Indexing on a Separate Drive.) Setting index options, such as Case Sensitive, Sounds Like, and Word Stemming, can also enhance how well an index can be searched. For more information, see "Creating a new index and choosing options" on page 272.

10 Exit or quit Catalog after “Index Build Successful” appears in the Catalog message window.

Structuring PDF collections for indexing

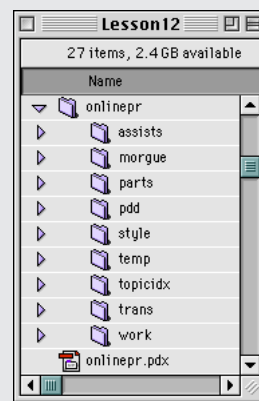
When you define and build an index, Catalog creates an index folder in which it puts an index definition file and a support folder. The index definition file has the same name as the index folder but has a .pdx extension. The folder has the same name as the PDX file and contains related folders that are automatically generated by Catalog.

Catalog places the PDX file (onlinepr.pdx in this example) and the support folder (onlinepr in this example) in the folder that contains the indexed document collection. The following guidelines apply:

- The entire index—both the PDX file and the support folder—must be located inside a single folder.
- The indexed documents must reside on a single disk drive or network server volume, and the index must be on the same drive or volume as the indexed documents (Windows).

Consider creating a separate PDF file for each chapter or section of a document. When you separate a document into parts and then search it, search performance is optimized.


—From the online Adobe Acrobat User Guide, Chapter 11



Searching an index

Now you'll use the Search command in Acrobat to perform searches of the PDF documents you just indexed with Catalog. You'll also use the Search command to limit and expand the definition of the term for which you are searching.

By searching a full-text index, you can quickly search a collection of PDF documents; in contrast, the Find command works only with a single PDF document and reads every word on every page, a much slower process.

 For information on the Find command, see “Finding words in PDF documents” in Chapter 1 of the online Adobe Acrobat User Guide.

1 Start Acrobat.

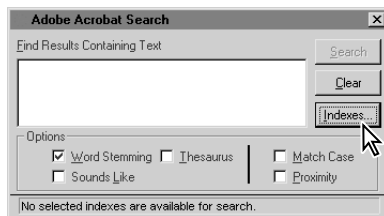
2 Begin the search by using the Search button () in the command bar:

- In Windows, choose Window > Show Command Bar if the command bar is not visible. Then click the Search button. In Windows, the command bar is visible whenever Acrobat is open and Window > Show Command Bar is selected.
- In Mac OS, choose File > Open. Select Basics.pdf in the Hawaii folder, located inside the Lesson11 folder, and click Open. Then click the Search button. In Mac OS, a document must be open in Acrobat for the command bar to appear. (You open the Basics.pdf document only to view the command bar, not as part of the search.)

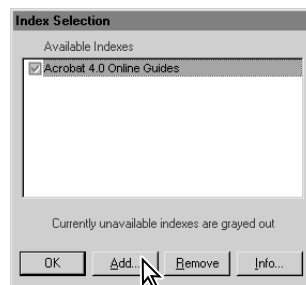
The Adobe Acrobat Search palette appears. First you'll select an index to search.

3 Click Indexes to display the Index Selection dialog box.

4 Click Add.



Adobe Acrobat Search palette

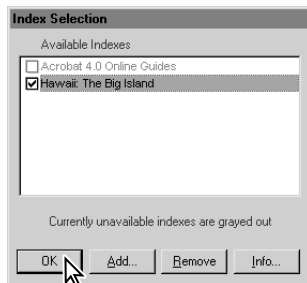


Index Selection dialog box lists available indexes

5 Select the Hawaii folder, located inside the Lesson11 folder, and click Open.

6 Select Guide.pdx, located inside the Hawaii folder, and click Open. The Hawaii: The Big Island index now appears under the list of available indexes.

7 Deselect any other indexes in the list by clicking the boxes next to them to clear the checkmark. Then click OK.

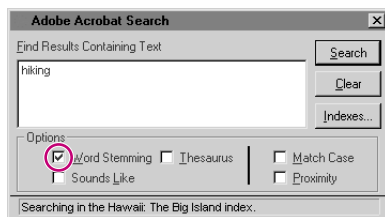


Acrobat searches only the selected index or indexes in the Available Indexes list. Here, you'll search only for entries in the Hawaii index.

To find information contained in the Hawaii index, you enter a word or phrase representing the desired topic.

8 In the Find Results Containing Text box, enter **hiking**.

9 For Options, select Word Stemming, and deselect all other search options.



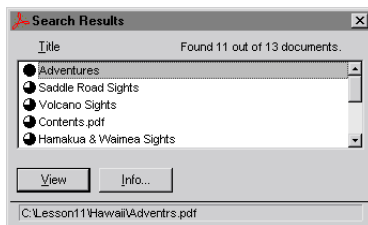
The Word Stemming option tells Search to look also for words that share the same word stem as “hiking,” such as “hike” or “hiked.”

[?] For detailed information on each search option, see “Setting the search options” in Chapter 12 of the online Adobe Acrobat User Guide.

10 Click Search.

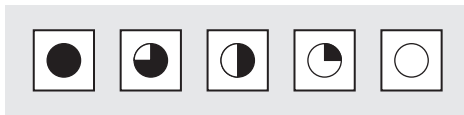
Looking at the Search Results list

The Search Results palette lists the documents that contain the word or words you searched for. The palette also displays how many documents were searched and how many were found to contain the words. In this example, 13 documents were searched, with 11 containing variations of the word “hiking.”




The documents are listed by relative ranking—documents with solid circles have more occurrences of the search words relative to the other documents in the list.

Acrobat Search uses five icons to indicate a document’s relevance ranking:

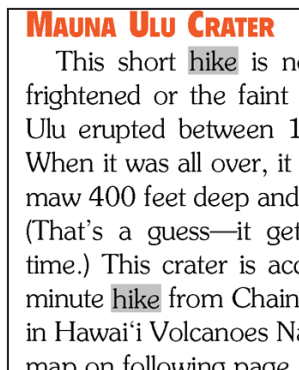


Relevance ranking from most (solid circle) to least (empty circle)

If needed, use the scroll bar or resize the Search Results palette to view the entire list. You can open any of the documents in the list and view the highlighted search words.

1 Select the Adventures entry at the top of the list, and click View to display the corresponding document. If needed, click the Fit in Window button () to view the entire document page.

Notice the highlighted occurrences of the word “hike.” Because the Adventures document contains many occurrences of the searched word, it is ranked near the top of the Search Results list. If desired, use the zoom-in tool (🔍) to magnify a section in which “hike” is highlighted.



- 2 Click the Search Results button (🔍) to redisplay the Search Results list. If needed, drag the Search Results list by its title bar to reposition the list so that it doesn't obscure the document.
- 3 Scroll to the bottom of the list, and double-click Kohala Sights to open that document. If needed, click the Fit in Window button to view the highlighted occurrence of “hiking.” The Kohala Sights document is ranked at the bottom of the Search Results list because it has fewer occurrences of the searched word than the other files in the list.
- 4 Choose File > Close to close the Kohala Sights document.

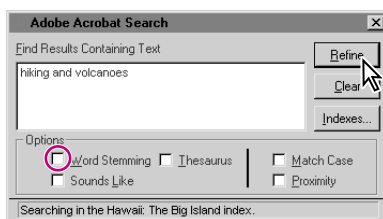
Narrowing the search

To make searching more effective, you should define your search criteria as much as possible. So far, you have found 11 documents that contain some information about hiking on the island of Hawaii. Now you'll refine your search to list only those documents that contain information about hiking near volcanoes.

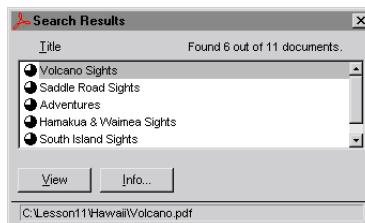
- 1 Click the Search button.
- 2 In the Find Results Containing Text box, after “hiking,” enter the words **and volcanoes**.
- 3 For Options, deselect Word Stemming to search only for the specific words you entered.

You are about to do a refined search. A refined search tells Search to look only at the documents in your current results list and to apply the new search criteria (instead of searching the index completely from scratch).

4 Hold down Ctrl (Windows) or Option (Mac OS) to change the Search button in the dialog box to Refine, and click Refine.



Word Stemming option deselected



Results of refined search


Notice that 6 documents out of the 11 in the last Search Results list meet the new search criteria. Using very specific search criteria helps you quickly identify which documents, if any, contain the information for which you are looking.

Also note that you've eliminated five documents with this search, but that the ranking of the remaining six documents is very close.

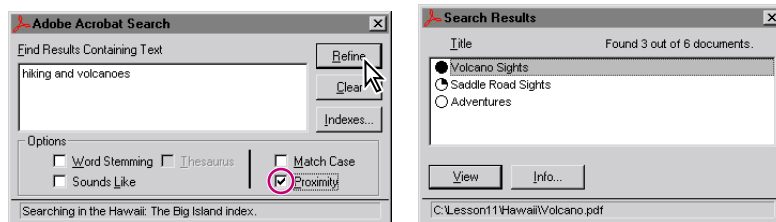
5 Close the Search Results palette.

Affecting the ranking order

You need to know which, if any, of the six listed documents is the best candidate for information about hiking near volcanoes. You can affect the relative ranking order by using any of the options displayed in the Adobe Acrobat Search dialog box.

- 1 Click the Search button () in the command bar.
- 2 For Options, select Proximity. This option finds documents in which the search words occur within three pages of each other. If the words are more than three pages apart, the search criteria aren't met and the document won't be listed in the Search Results window.
- 3 Hold down Ctrl (Windows) or Option (Mac OS) to make the Search button change to Refine, and click Refine.

Notice that the Search Results list has narrowed the field to three documents with very different relative rankings.



Refining search to narrow results

Result

4 Select Volcano Sights in the list and click View, or double-click Volcano Sights. To view the whole document page, click the Fit in Window button.

Notice that nearby occurrences of “hiking” and “volcano” are highlighted. You can zoom in on the highlighted text to read the information.

5 Choose File > Close to close the document.


Using Document Info fields to search

In addition to conducting a search based on specific words, you can search a collection of documents using any one of the Document Info fields. In this section, you’ll search for documents that have a specific entry in their Subject fields.

First, you’ll make sure that the Adobe Acrobat Search dialog box displays Document Info fields.

1 Choose File > Preferences > Search to display the Acrobat Search Preferences dialog box.

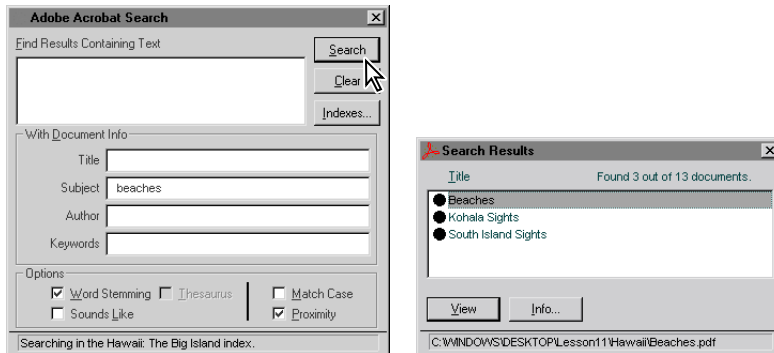
2 For Query, select Show Fields, and click OK.

3 Click the Search button () in the command bar.

The Adobe Acrobat Search dialog box now displays Document Info fields in which you can enter additional search information. For example, when creating an index for a document, you can enter relevant information in the Title, Subject, and Keywords fields to make a search more efficient. Users can then search for this specific information.


4 If needed, click Clear to clear the text and Document Info fields.



5 For Subject, enter **beaches**, and click Search. Three documents meet the search criteria.



Entering search criteria

Search results

6 Double-click **Beaches** to open that document. To view the whole page, click the Fit in Window button ()

7 Click the Next Highlight () button to open the next document in the Search Results list. Once you click the Next Highlight button, the Previous Highlight () button becomes available. You use the Next Highlight and Previous Highlight buttons to browse through a Search Results list.

***Note:** Alternatively, use the Next and Previous Documents commands in the Edit > Search menu to browse the search results.*


8 Choose File > Close to close the document.

Searching with Boolean expressions

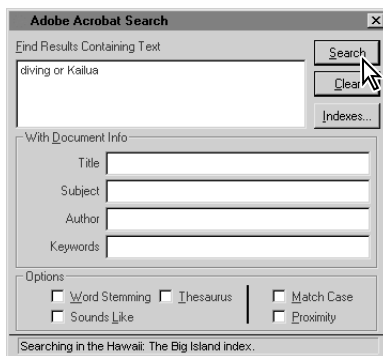
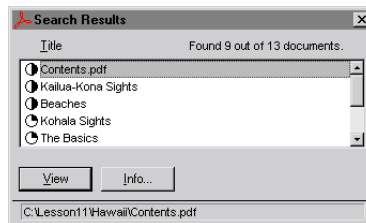
You can use AND, OR, and NOT operators to build a logical expression (called a *Boolean expression*) that searches for words in a specific relation to each other. For example, earlier in this lesson you used the AND operator to build an expression that searched for occurrences of both “hiking” and “volcanoes” in the same document.

Using the OR operator

Use OR to find documents containing any of two or more search terms. Maybe you aren’t sure what you want to do one day on your vacation—you could go diving or to Kailua.

- 1 Click the Search button () in the command bar.
- 2 Click Clear to clear the text and Document Info fields. Deselect all search options.

- 3 In the Find Results Containing Text box, enter **diving or Kailua**.
- 4 Click Search.


*OR Boolean expression**Results*

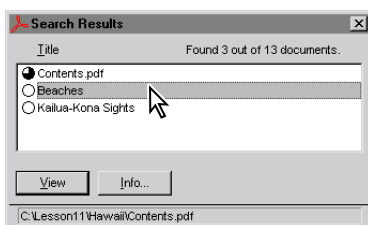
The Search Results palette lists nine documents that contain information about either diving or Kailua. You can open any of the documents to find information about one subject or the other, but not necessarily both.

- 5 Close the Search Results palette.


Using the AND operator

If you plan it right, you might be able to go diving and see Kailua in one day trip. Use AND to find documents containing two or more search terms.

- 1 Click the Search button () in the command bar.
- 2 Click Clear. Under Find Results Containing Text, enter **diving and Kailua**.
- 3 Click Search. Three documents contain information about both diving and Kailua.
- 4 Double-click Beaches to open that document. Zoom in on the bottom right quarter of the displayed pages to learn about a small cove near Kailua that is excellent for scuba diving.



You can also search with the NOT Boolean operator, combine Boolean operators, and search with wildcard characters.

 For more information on searching with Boolean operators, see “Tips on defining search queries” in Chapter 12 of the online Adobe Acrobat User Guide.

5 Choose File > Close and close all open files (Windows) or close all open files except the Basics.pdf file (Mac OS).

Searching on the Web

An index created by Catalog is not searchable over the Web or a company intranet, but your documents can be indexed by a Web search engine that supports indexing PDF documents. You must first set up the search-engine software that can search PDF files on your Web server. Users can then run a search from your Web page.

Many companies develop search engines that can automatically and continuously update a search index of both PDF and HTML documents on a Web server. For a current list of these companies, see “Searching PDF Files on the Web” on the Acrobat Web page at <http://www.adobe.com/acrobat/moreinfo>, and click a specific link.

Choosing Catalog options

You can change the Catalog default options to minimize the size of an index as much as possible, for example, if you plan to post an index (and its associated files) on a network file server where space is at a premium.

In Windows, the defaults are fixed. You can change them for a particular definition, but not permanently. In Mac OS, you can change the defaults for most of the options in the Index Defaults group of preferences.

Creating a new index and choosing options

In this section, you’ll create another index of the documents contained in the Hawaii folders and choose Catalog options to reduce the size of that index.

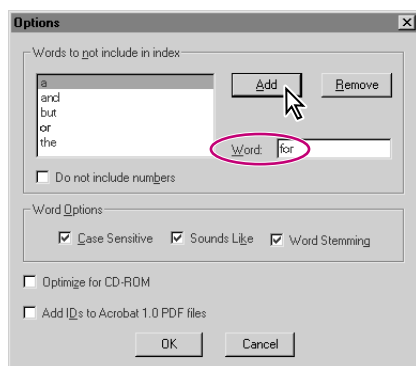
- 1** Start Acrobat Catalog.
- 2** Choose Index > New (Windows) or File > New (Mac OS).
- 3** For Index Title, enter **The Island**.
- 4** Click Options.

The Options dialog box lets you change Acrobat Catalog defaults for a particular index definition. You can exclude specified terms (called *stopwords*) and numbers, and turn off the Match Case, Sounds Like, and Word Stemming features. If the collection contains PDF files created by version 1.0 of Acrobat PDF Writer or Acrobat Distiller, you can select the Add IDs to Acrobat 1.0 PDF Files option to add unique IDs to the file to assist Search in locating a document.

Use the Words To Not Include In Index section to enter stopwords—words that you exclude from an index to minimize its file size. Stopwords usually are words that would not be entered as search terms, including articles such as “the” and “a”, conjunctions such as “but” and “or”, and prepositions such as “for” and “by.”

Excluding stopwords from an index makes the index typically 10% to 15% smaller. However, searches won’t find phrases (that is, words enclosed in quotation marks) that contain the stopwords. To help users, you should list the stopwords in the index description.

5 For Word, enter the following words, clicking Add after each entry: **the, The, and, And, a, A, but, But, or, Or, for, For, by, By.**




Note: If stopwords include words that are also used in Boolean expressions, such as “and” and “or,” the Search feature won’t be able to perform logical searches with those Boolean operators.

You can add up to 500 stopwords to an index. Stopwords can be up to 24 characters long. If you want to stop a word completely, you must enter all capitalization possibilities such as “the” and “The.”

6 Select Do Not Include Numbers. This option excludes numbers such as phone numbers, part numbers, and address numbers from an index.

Excluding numbers from an index about Hawaii is appropriate because users of the index will not be looking for specific numbers. Excluding numbers from an index of an auto parts manual, where it is likely people will be searching for numbers, is not appropriate.

7 Select all of the Word Options. These options let you use some of the options in the Search dialog box.

 For more information on Word options, see “Setting the search options” in Chapter 12 of the online Adobe Acrobat User Guide.

8 Deselect Optimize for CD-ROM.

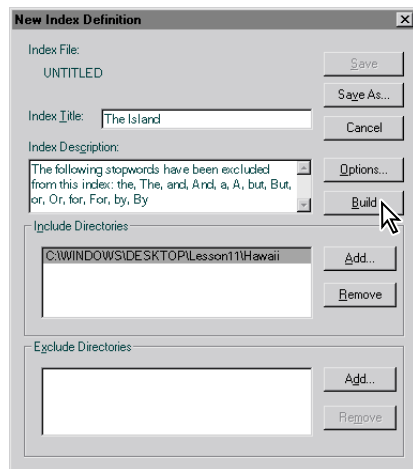
Intended for indexes placed on a CD, this option arranges index files for the fastest possible access. Additionally, this option makes it easier to modify Document Info fields or security settings after you have indexed a document by bypassing a default alert message.

9 Deselect Add IDs to Acrobat 1.0 PDF Files, because this collection doesn’t contain any 1.0 PDF files. Click OK.

10 In the Index Definition dialog box next to the Include Directories box, click Add. Select the Hawaii folder, located in the Lesson11 folder, and then click OK (Windows) or click Select “Hawaii” (Mac OS).

11 In the Index Description text box, add information on what stopwords were included. (For example, we entered: “The following stopwords have been excluded from this index: the, The, and, And, a, A, but, But, or, Or, for, For, by, By.”)

12 Click Build. Open the Hawaii folder so that you can save the index within that folder. Name the index **Island.pdx**, and click Save.



13 Exit or quit Catalog.

Comparing index sizes

Indexes consist of an index definition file (PDX) and a folder that contains the supporting data files that are needed to return search results. This folder must travel with the PDX file at all times. An index definition file also contains relative paths between the PDX file and the folders containing the indexed documents. These relative paths must also always be maintained to complete successful searches.

When you set options in Catalog to reduce the size of your index, you are actually reducing the size of the supporting data files and not the index definition file itself. Now you'll compare the sizes of the folders that hold the supporting data files to see how the folder size is reduced.

1 From the desktop, open the Hawaii folder, located in the Lesson11 folder. Among the files, you should see two folders, named Guide and Island.

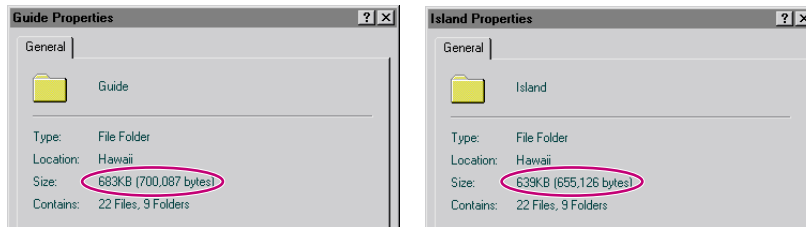
2 Display the folder size:

- In Windows, right-click the Guide folder and choose Properties.
- In Mac OS, click once to select the Guide folder, and choose File > Get Info.

3 Write down the folder's size. In Mac OS, close the Get Info window.

4 Repeat steps 2 and 3 for the Island folder.

As you can see, the Island folder is smaller than the Guide folder. If these indexes were larger, you would see an even greater difference in the size of the folders.



Guide folder size

Island folder size

Tips for updating indexes

You must update an index if documents are added to or removed from the collection, or if the hierarchy of the indexed folders has changed.

You should also consider updating an index when documents in the indexed document collection have changed, or data values for the new field are added because a new Document Info field has been defined.


You can reduce the index update time by following these guidelines:

- Don't support the Sounds Like, Case Sensitive, or Word Stemming search options.
- Use stopwords and exclude numbers.
- Install Acrobat Catalog on the system where the indexed documents are stored. If the program and documents are on different systems and it is feasible to move the documents temporarily, move them to the Catalog system for updating, then move them back.

—From the online Adobe Acrobat User Guide, Chapter 11

Searching for a phrase

By using the Search button in Acrobat or Reader, you can search for a phrase that contains two or more words, such as *Marriage license*. But when you search for a phrase containing words that might be confused with Boolean operators such as *and*, *or*, and *not*, or parentheses, you must enclose the phrase in quotation marks.

- 1 In Acrobat, click the Search button () in the command bar.
- 2 Click Indexes.

3 If the *Hawaii: The Big Island* index created in this lesson is not in the available indexes list, click Add. Select Guide.pdx, located in the Hawaii folder inside the Lesson11 folder, and click Open.

Now you'll add the index created in this lesson with stopwords.

4 Click Add. Select Island.pdx, located in the Hawaii folder inside the Lesson11 folder, and click Open. *Hawaii: The Big Island* and *The Island* are listed as available indexes.

5 Deselect *The Island* so that the index remains in the list, but won't be searched. Click OK.



6 In the Find Results Containing Text box, enter **Mauna Loa and Mauna Kea**; then click Search. (These are the two biggest volcanoes on Hawaii.)

The Search Results palette lists eight documents that contain information about the two volcanoes. The search engine returned all documents that contained both Mauna Loa and Mauna Kea to the Search Results list, even if both names weren't in the same sentence. But what if you remember seeing something about the two volcanoes in the same sentence and want to find the phrase *Mauna Loa and Mauna Kea*? If that is the case, you need to add quotation marks.

7 Click the Search button in the command bar.


8 In the Find Results Containing Text box enter **"Mauna Loa and Mauna Kea"** (with quotation marks) and click Search.

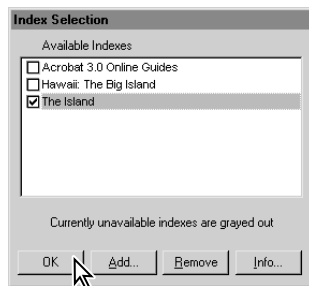
Now the Search Results list returns two documents that contain the phrase "Mauna Loa and Mauna Kea."

9 Double-click Saddle Road Sights to view the information. Close the file when you have finished viewing it.

Seeing the effects of stopwords

As mentioned earlier, excluding certain words from an index results in a smaller index. But searching for phrases that contain stopwords will fail and index users won't be able to find phrases that contain the stopwords.

- 1 Click the Search button () in the command bar.
- 2 In the Adobe Acrobat Search dialog box, click Indexes.
- 3 Deselect *Hawaii: The Big Island* so that the index remains in the list, but will not be searched.
- 4 Select *The Island*. Click OK.



- 5 In the Find Results Containing Text box enter **Mauna Loa and Mauna Kea** (no quotation marks) and click Search. The Search Results palette lists eight documents that contain information about these two volcanoes.
- 6 Click the Search button in the command bar.
- 7 Now enter “**Mauna Loa and Mauna Kea**” and click Search.

Remember that you excluded *and* from the index as a stopword. As a result, the current search found no documents that matched your query. The use of stopwords prevents you from locating phrases (as defined by quotation marks) containing words included as stopwords, even though these phrases exist in the document collection.

- 8 Click OK to close the alert box.
- 9 Exit or quit Acrobat.

This completes the lesson. You've learned how to build an index using Acrobat Catalog and how to perform various searches. For additional practice in building indexes from a series of PDF files, see Lesson 13, “Distributing PDF Documents.”

Exploring on your own

If you're including PDF files on a CD, Adobe recommends that you also include an index to help your users access your information with ease and efficiency. You can practice building an index for a CD from PDF files.

For additional practice, you can update an index (called scheduling an index build). Scheduling index builds or updates lets you keep your indexes up to date and use system resources when they are usually not needed for other tasks. Catalog must be running to execute the build or update.

1 Start Catalog.

2 Choose Index > Schedule.

3 For each index you want to build, click Add, and select the name of the index-definition (PDX) file. (If you have about 6 MB of free RAM, you can add up to 128 indexes.)

For example, select Island.pdx in the Hawaii folder, located inside the Lesson11 folder.

4 Click Open.

5 Select a build time: Every, Once, or Continuously.

- If you select Once or Continuously, the index begins building or updating immediately.
- If you select Continuously, Catalog updates the indexes in the order they are listed in the Indices to Build list.
- If you select Every, enter a numeric time interval, and then choose Minutes, Hours, or Days from the menu. If you want to delay processing the selected indexes until a specified time, select Starting At, and use the Up and Down Arrow keys to select the time. Select Every for updates at regular intervals. Then enter a time interval in the format hh:mm:ss (for hours:minutes:seconds). If you want to delay processing the selected indexes until a specified time, select Starting At and use the arrow keys or type in a number to select the time.

Catalog updates are incremental, to minimize updating time and permit searching to go on uninterrupted during updates.

6 Click Start.

Note: To save the schedule information and enter Scheduled Build mode later, click Save in the Scheduled Builds dialog box.

7 Click Stop to halt the update process.

Review questions

- 1 How can you distinguish an index file from a PDF file?
- 2 What does the Search Results palette display?
- 3 How does the Word Stemming option expand your search results?
- 4 What key do you press to do a refined search?
- 5 Why should you enter Document Info field information in all your PDF documents?
- 6 Give two examples of Boolean expressions and what they do.
- 7 What are stopwords?
- 8 How can an index user find out which, if any, stopwords were excluded from an index?
- 9 When would you not include numbers in an index?
- 10 Can you schedule an index to build every 30 minutes?
- 11 When should you purge an index?
- 12 What is the difference between searching for a phrase with quotation marks and without quotation marks?

Review answers

- 1 In Acrobat, the file extension PDX identifies a file as an Acrobat index.
- 2 The Search Results palette lists the documents that contain the word or words you searched for, and displays how many documents were searched and how many were found to contain the words.
- 3 The Word Stemming option tells Search to look also for words that share the same word stem. For example, the word “hiking” also includes the word stems “hike” or “hiked.”
- 4 A refined search is one in which Search looks only at the documents in your current results list and applies the new search criteria (instead of searching the index completely from scratch). To do a refined search, hold down Ctrl (Windows) or Option (Mac OS) to change the Search button in the dialog box to Refine.

5 To give users more ways to search for information in a PDF file, you should fill in the Title, Subject, and Keywords text boxes in the Document Info dialog box with the relevant information. Users can search for this specific information and use the title, subject, and keywords as they would index terms to refine their searches.

6 AND, NOT, and OR are Boolean expressions used in searches, as follows:

- AND combines two or more search terms.
- OR searches for any of two or more search terms.
- NOT excludes terms from searches.

7 Stopwords are words that you exclude from an index to minimize the size of the index file or reduce it by 10% to 15%. Stopwords are usually words that would not be entered as search terms; for example, articles such as “the” and “a,” conjunctions such as “but” and “or,” and prepositions such as “for” and “by.”

8 Listing the stopwords in the index description helps users find out what stopwords were excluded from the index.

9 You should exclude numbers from an index that doesn’t pertain specifically to numbered information, such as in a guide book or descriptive text. But you should include numbers in an index if users would be searching for numbers, for example, for technical specifications, addresses, part numbers, or phone numbers.

10 You can schedule an index to build at regular intervals, including every 30 minutes, or once only or continuously.

11 To compensate for incremental Catalog updates that increase the size of the index, you should purge and rebuild an index periodically to reclaim disk space and speed up searches.

12 Searching for a phrase enclosed in quotation marks finds the verbatim phrase. Searching for a phrase without quotation marks finds the terms within close proximity in the document.