How to Find Specific Words and Phrases in Searchable Books/Articles in the Puritan Hard Drive (PHD) PDF Collection

Introduction

The Puritan Hard Drive PDF collection contains more than 1,500 books/articles that have been rendered searchable through the application of Optical Character Recognition (OCR) software. To greatly facilitate searching these documents, each one has an embedded full text index, which is not visible to the reader, and a specialized index (from a base of 5,200 key words and phrases) that has been appended to the document itself and can be directly read by the user. The appended indices have also been incorporated into the PHD KnowledgeBase software and, along with such attributes as author, title, description, and subject category, can be used to locate particular documents of interest to the user. How to use this keyword/phrase search within the PHD KnowledgeBase is explained in depth in the forthcoming User's Guide and video tutorials.

Once a document has been loaded into Adobe Reader or Adobe Acrobat (Standard or Pro Edition), the embedded full text index is automatically accessed whenever the find or search command is executed within the application. The user doesn't need to apply any special settings to take advantage of the search speed enhancement provided by the embedded index.

Another feature has been included in the Puritan Hard Drive collection that is incredibly powerful, and potentially very useful to the user engaged in serious study and research. We have included a **Master Full Text Index** that aggregates all the indices of the individual documents and permits the user to enter a search query across the entire searchable PDF collection. This search will return a hierarchical listing of all the instances of the search word or phrase – in context – organized by the title of the document in which it is found. When a search result is clicked, the target document is automatically opened in the appropriate PDF application to the exact page where the search phrase is located and is highlighted for easy access. Depending on the number of instances returned from the search query and the speed of the particular computer on which the search is executed, the time it will take for the complete results listing to appear will vary from less than a minute to five or more minutes. The time it takes to execute the query is, of course, an infinitesimal fraction of the time it would take to manually execute the search in each individual document.

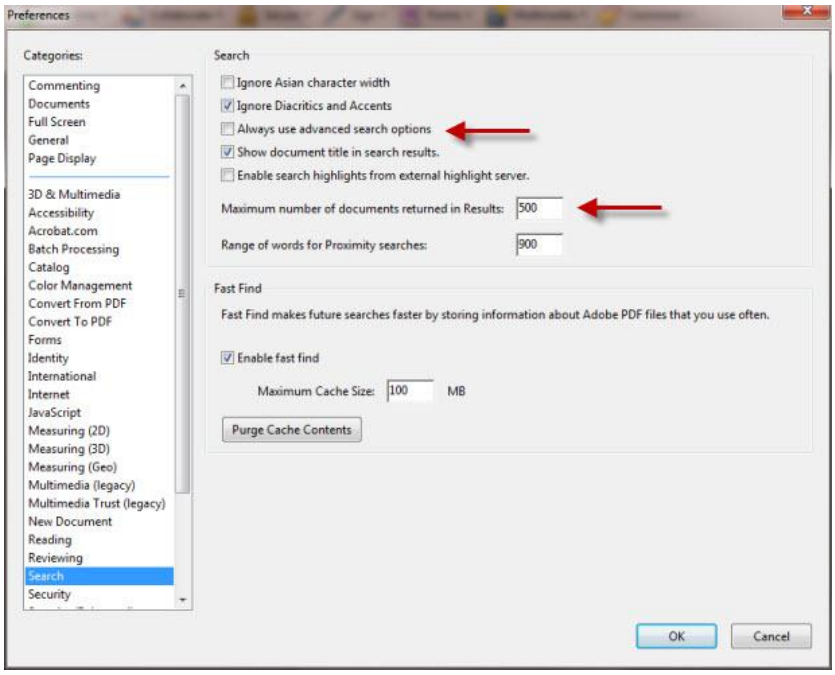
How to Use the Master Full Text Index

Unlike the embedded index, accessing the Master Full Text Index requires a brief setup by the user, which is explained in detail below. The first task is to enable "Advanced Search Options" to appear by default when executing a search and setting the maximum number of documents to be returned in the search results.

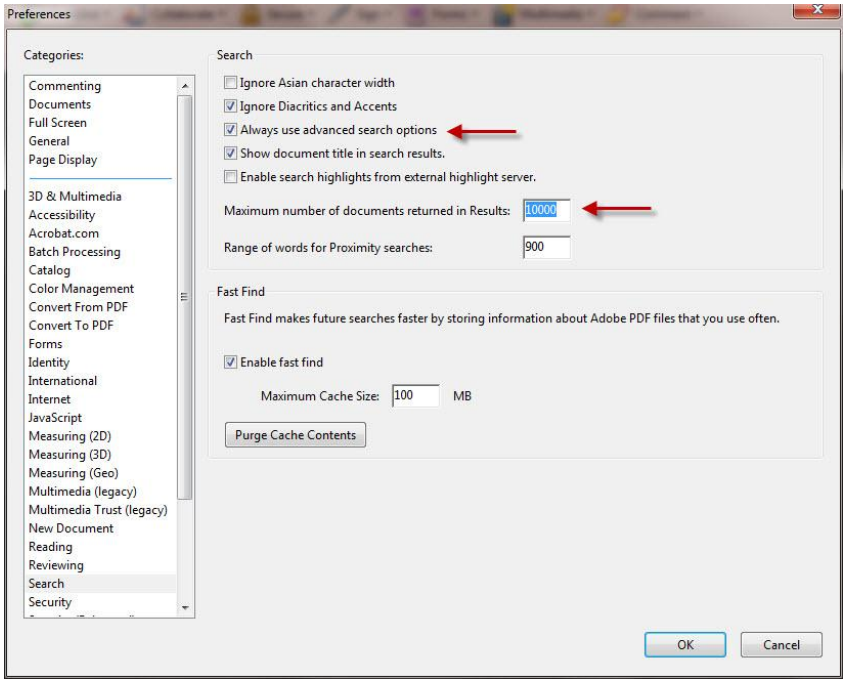
Setting Search Preferences on MS Windows Computers

(These instructions apply equally to Adobe Reader and Adobe Acrobat applications. Unless otherwise noted, all references to "Click" imply using the "Left Click" button on your mouse)

(1) From the main menu of your Adobe PDF application, click on the "Edit" option. From the drop down menu that appears, click on "Preferences".
(2) On the left hand side of the Preferences dialog box that pops up, click on the "Search" category,
(3) There are two settings we recommend changing (assuming the default values are currently set).
    a. "Always use advanced search options" – if it is not already checked, click on the check box to enable it. This will bring up the Advanced Search dialog box whenever the search command (from the main "Edit" menu) is executed.
    b. "Maximum number of documents returned in Results" – set this value to 10000 (that is the maximum value allowed; enter only numbers, do not enter a comma). This will ensure that all searchable documents in PDFs folder in the Puritan Hard Drive directory will be included in your search queries. Even though there are only (ONLY!) 1564 searchable PDFs currently distributed with the PHD, setting the value to 10000 will allow you to add additional PDFs as they are made available in the future without having to change the value in Preferences.
    c. The other options can be left at the default setting.
    d. Click on the "Ok" button to save your changes and close the "Preference" dialog window.



Preferences Dialog – with Default Values

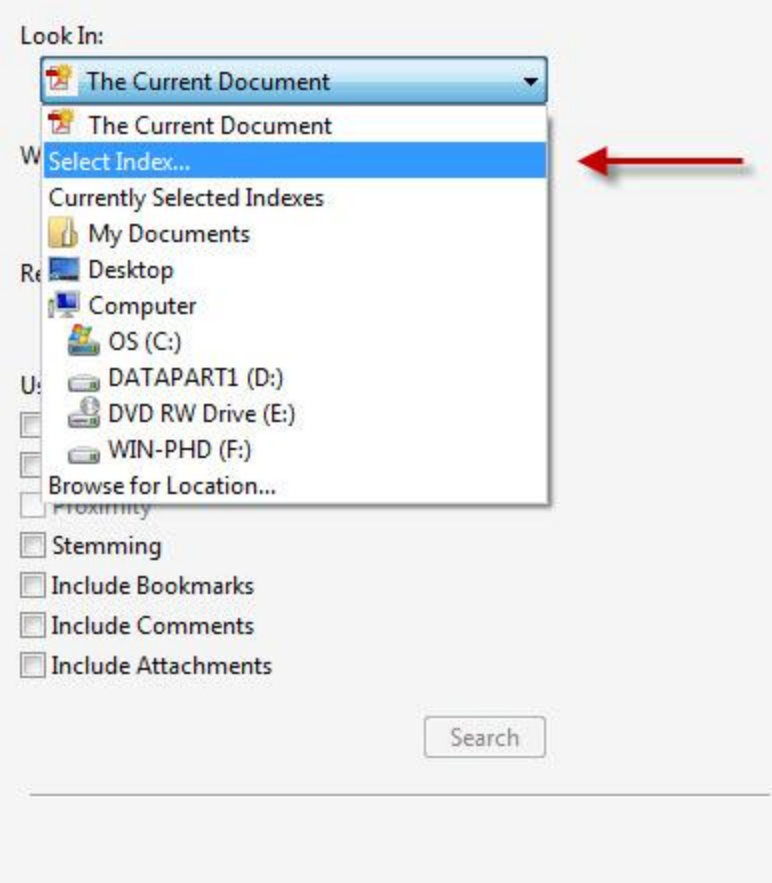


Preferences Dialog – with Recommended Values

Selecting the Master Full Text Index in the PHD Directory

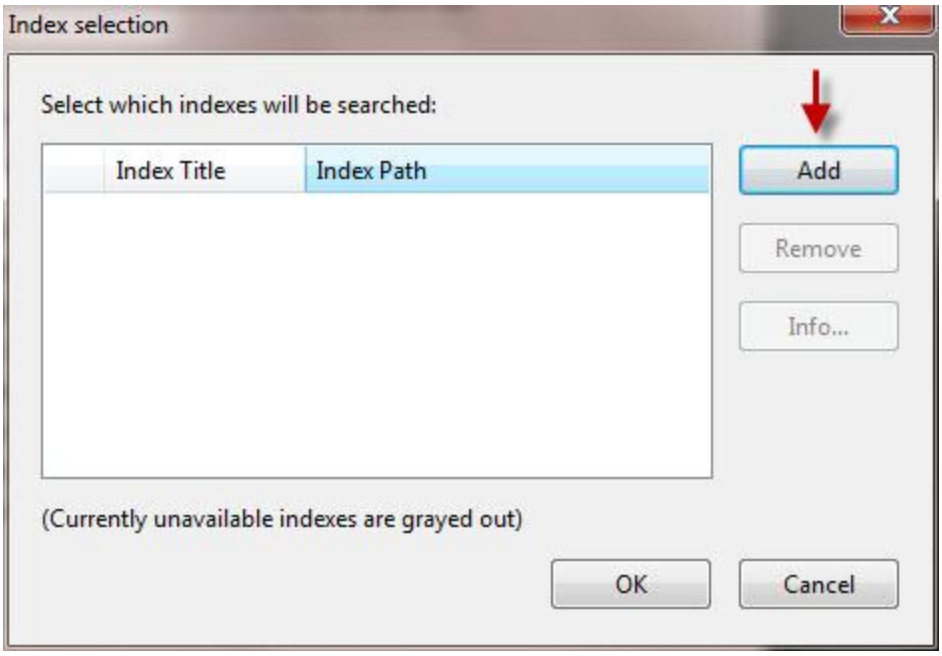
The next task is to tell the Adobe PDF application where to find the Master Full Text Index in the PHD directory. As long as you don't move the PHD directory to another location, this task only has to be done once. The application will store this information for future use. This task should be performed when there is no document loaded in the Adobe PDF application

(1) From the main menu of your Adobe PDF application, click on the "Edit" option. In the drop down menu that appear, click on the "Search" menu option.
(2) A standalone "Search" window will appear. The first item is called "Look in:" and immediately below it is a dropdown box that defaults to "The Current Document". Click on the down arrow on the right side of the drop down box. Highlight the "Select index…" option with the mouse cursor and click it. This will bring up another dialog box that will allow you to select the Master Full Text Index in your PHD directory.



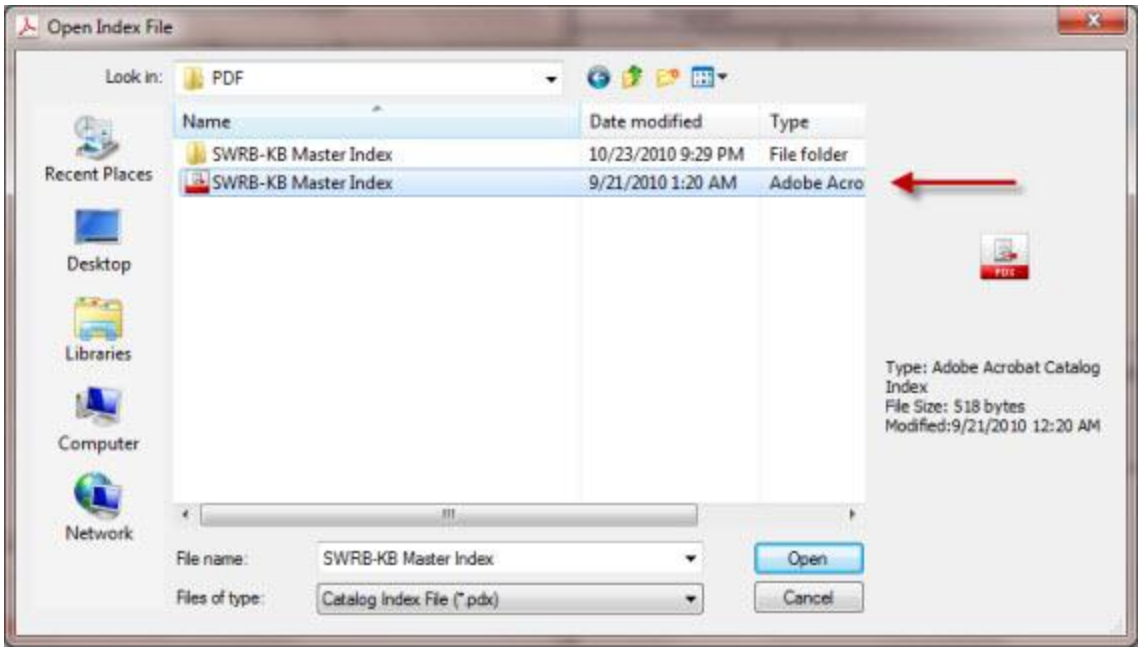
"Look In: " Dropdown Box

(3) Click on the "Add" button. This will launch a system file browser (Windows Explorer). For the purposes of these instructions, we will assume that you are using the PHD from the external USB drive that you received. If you have copied the PHD directory to an internal drive on your desktop or laptop computer, you will need to navigate the file browser to the top level "Puritan Hard Drive" directory. The remainder of the instructions will be the same wherever your directory is physically located.

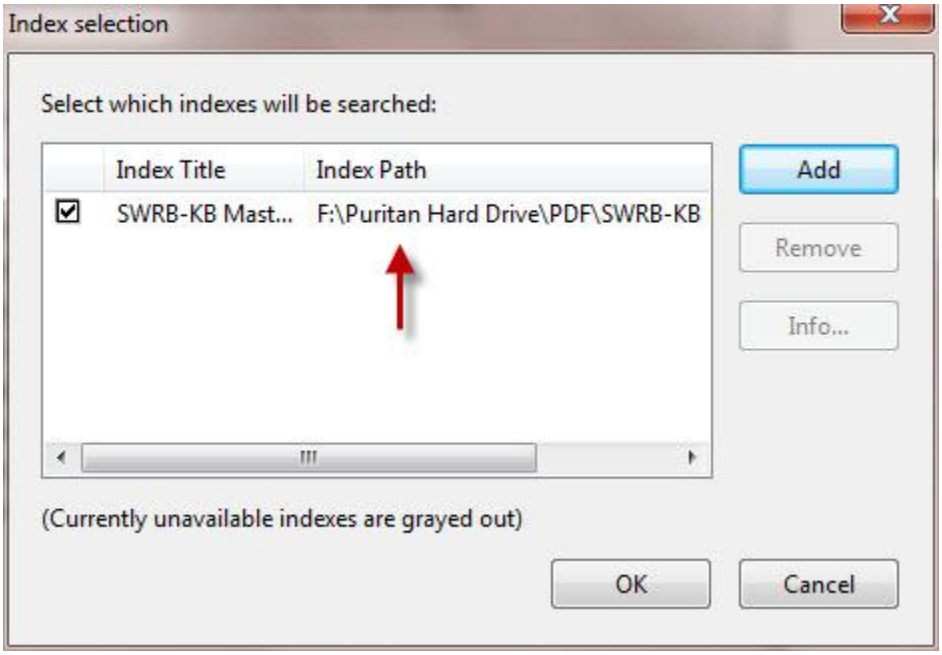


Index Selection Dialog Window

(4) To locate the Master Full Text Index in Windows Explorer file browser, click on the “Puritan Hard Drive” top level folder. Next, click on the “PDF” folder. Two folders should now appear in Explorer, a folder called “SWRB-KB Master Index” and a file with the same name. Click once on the file, not the folder and then click on the “Open” button. Explorer closes and the already opened “Index Selection” window regains focus. Click on the “Ok” button to accept the “SWRB-KB Master Index” just selected.

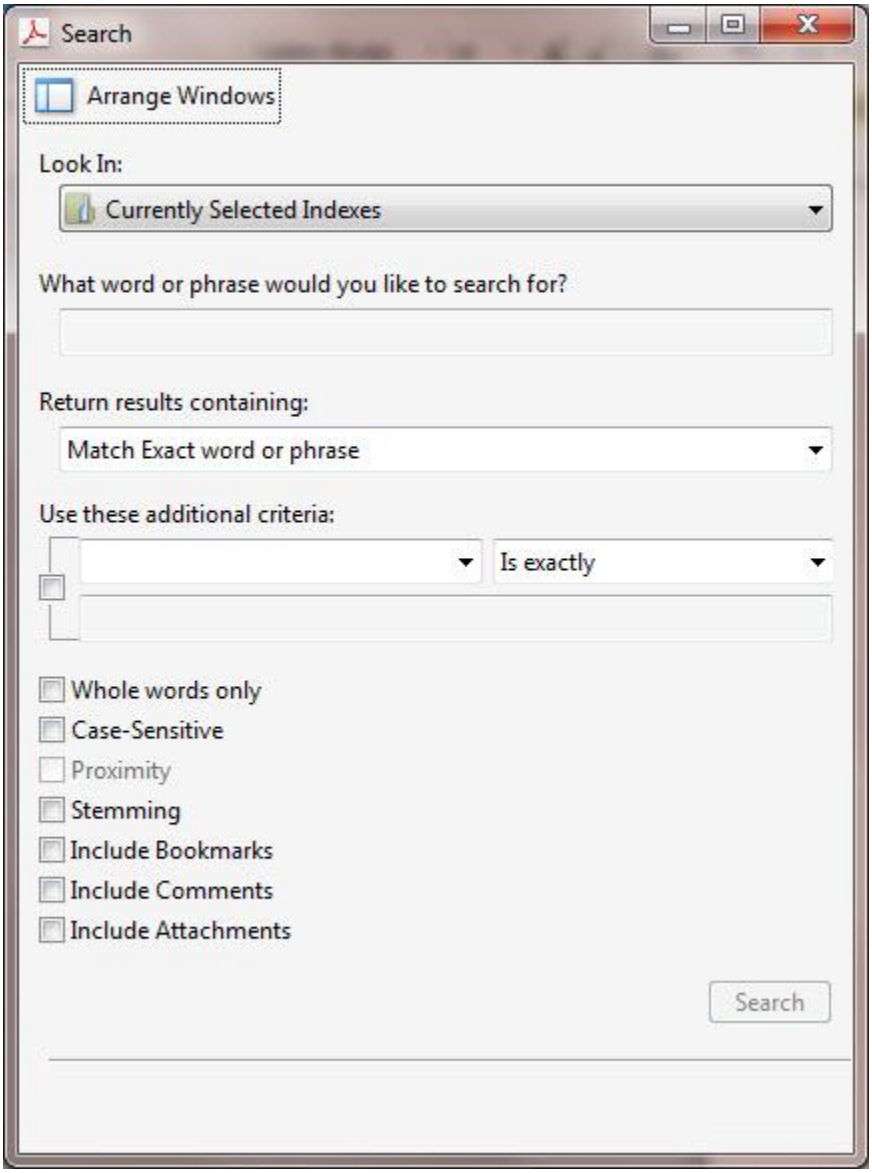


Open Index File Browser



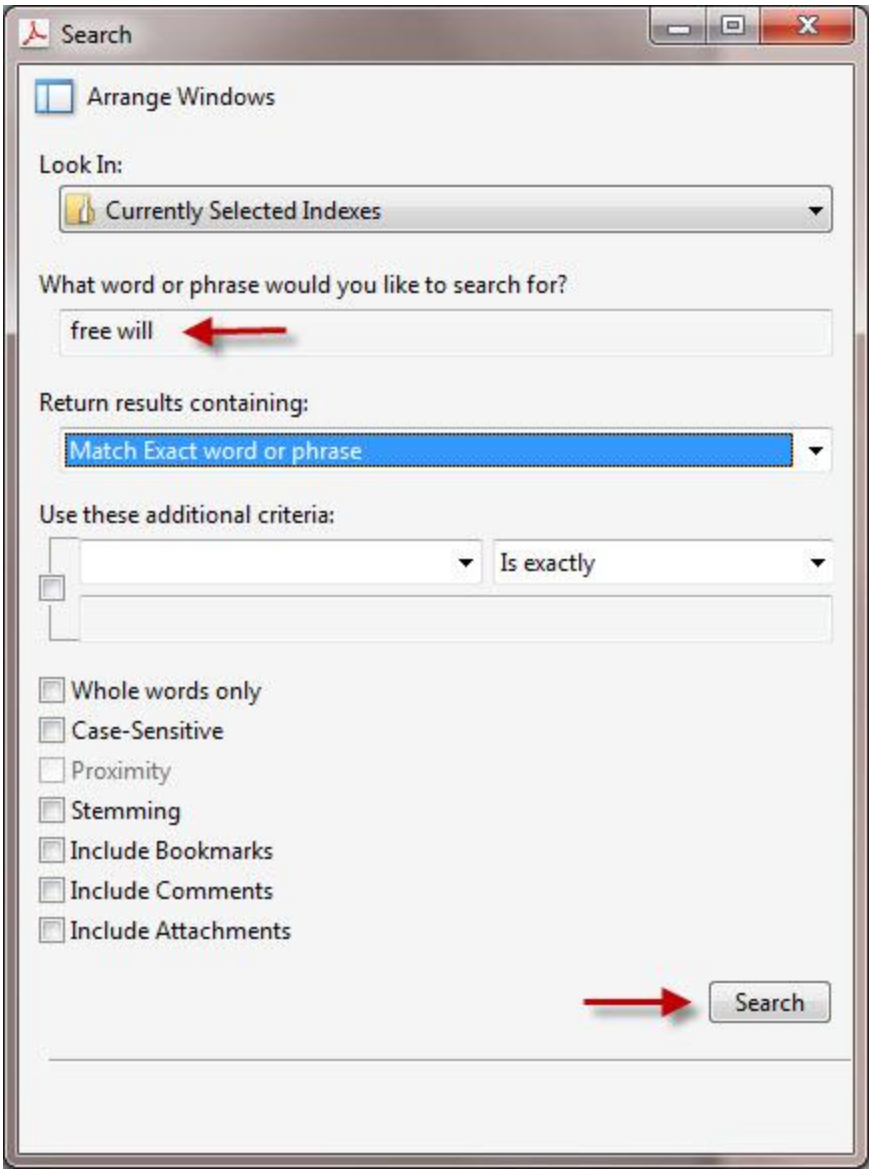
Index Select Dialog Window

(5) The “Search” window regains focus with the “Look In: “ dropdown box populated with “Currently Selected Indexes”.



Search Window with “Currently Selected Indexes” Option

(6) Now that the Adobe PDF application knows where the Master Full Text Index is located on your drive, you can enter a search word or phrase you are interested in locating in the searchable PDFs in the PHD collection. For example, suppose you want to find the phrase “free will” in the more than 1500 searchable PDFs. In the text box under the label “What word or phrase would you like to search for?”, enter the phrase “free will”.



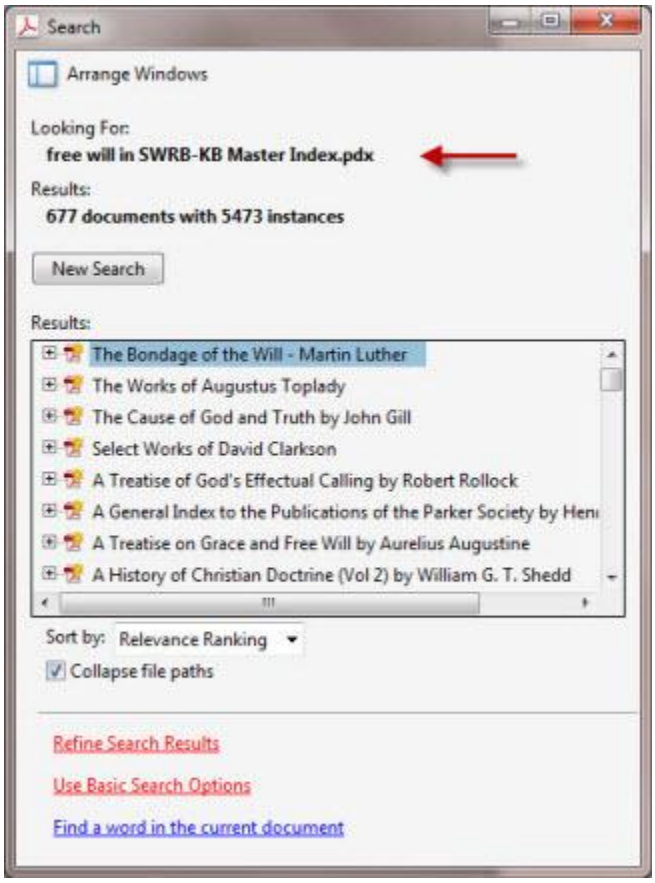
Entering "free will" in Search Text Box

(7) There are numerous options available in this dialog window to control the search process. For detailed information on these options, you should consult the documentation available in your Adobe PDF application. For this example, the application will match the exact phrase you entered into the search text box.

(8) Once you have entered the search phrase, "free will", click on the "Search" button to initiate the search process. How long this search process will take depends on a number of factors that can be categorized as computer dependent and content dependent. Computer dependent: processor

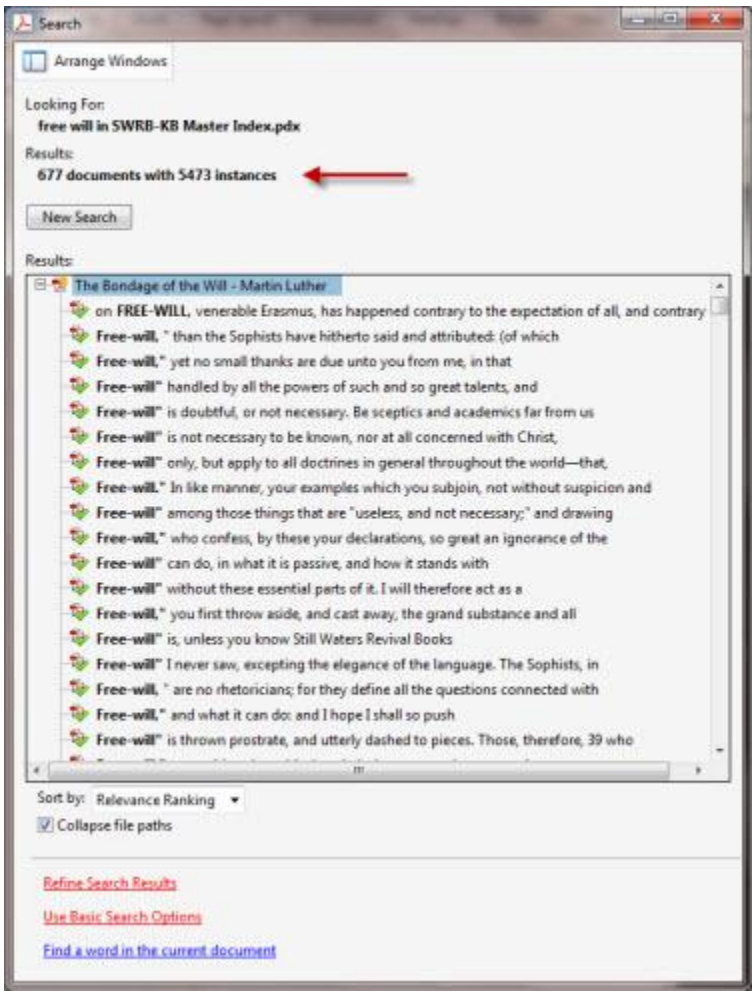
speed, available memory (RAM) and disk speed. Content dependent: the number of instances returned as search results. As a general rule, the less specific the search phrase, the more search results will be returned (potentially tens of thousands) and the longer the search process will take to complete. The Adobe PDF application will indicate that a search is in process and inform you when it is completed. It is not unusual for this process to take more than five minutes when returning thousands of results across hundreds of documents.

(9) The search for "free will" returns 5,473 results from 677 PDF documents. The results are organized by the title of the PDF document listed in order of relevance, i.e. number of results returned for each document (a minimum of one instance is required for the document to appear in the results list).



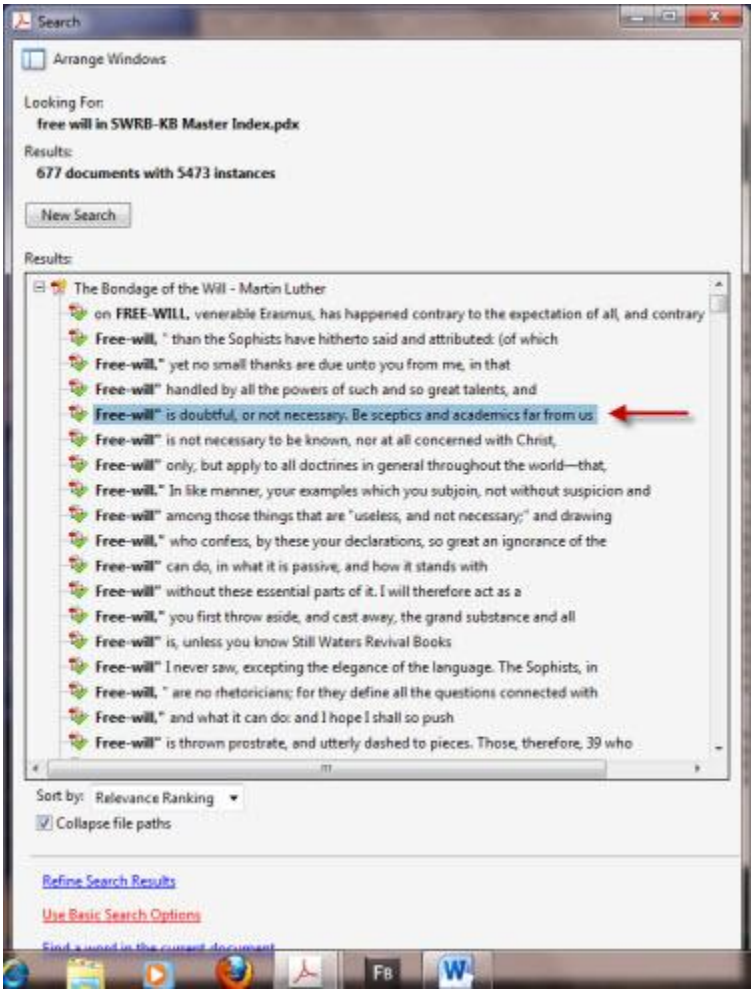
Search Results Organized by Document Title

(10) To access the individual search results within a document, click on the small plus sign (“+”) to expand the results listing. The listings will show the search phrase **in the context** in which it appears in the document.

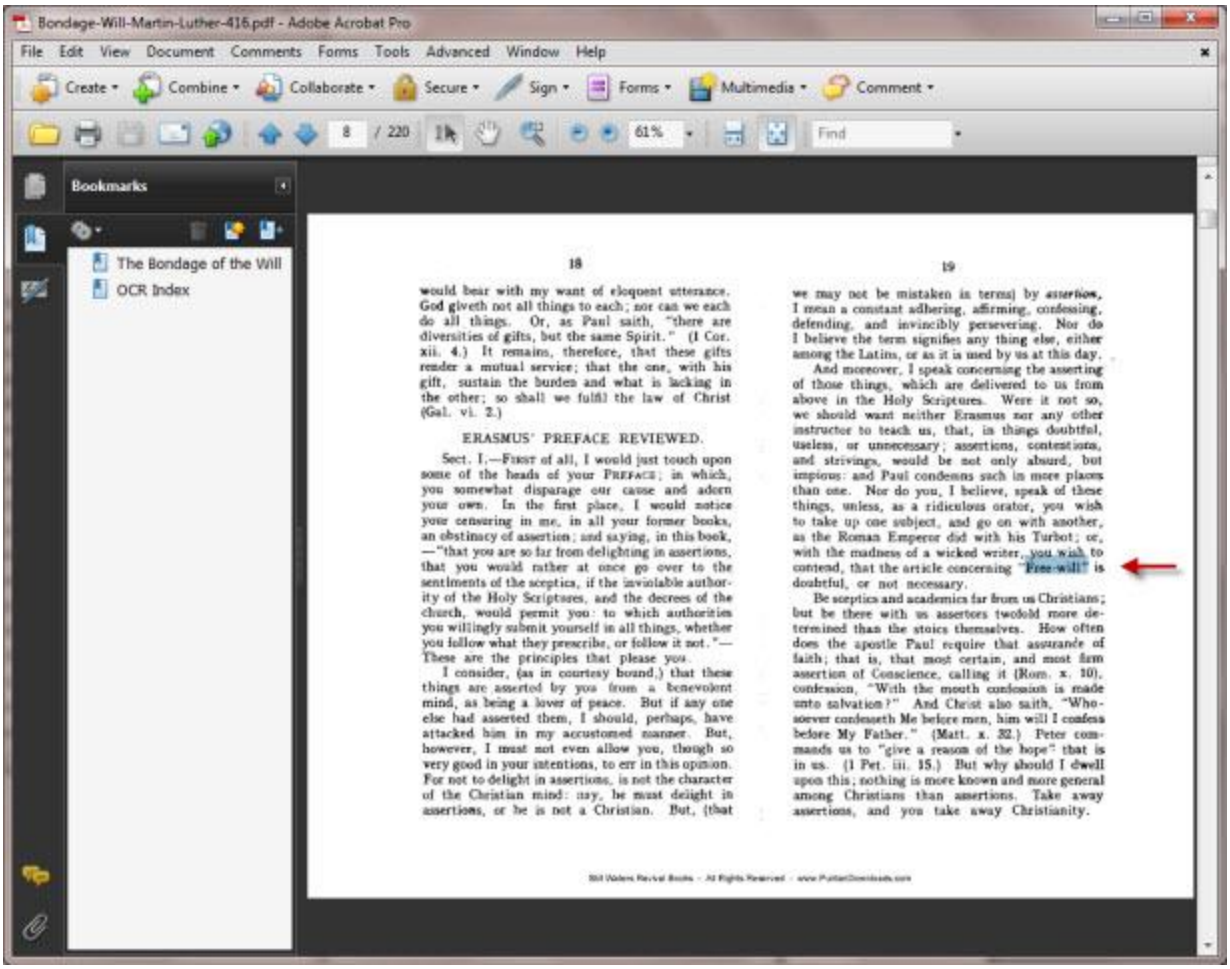


Individual Search Results Within “The Bondage of the Will”

(11) The individual search results are links that when clicked will open up the PDF document to the exact page where the search phrase is located, and the phrase will be highlighted for immediate access.

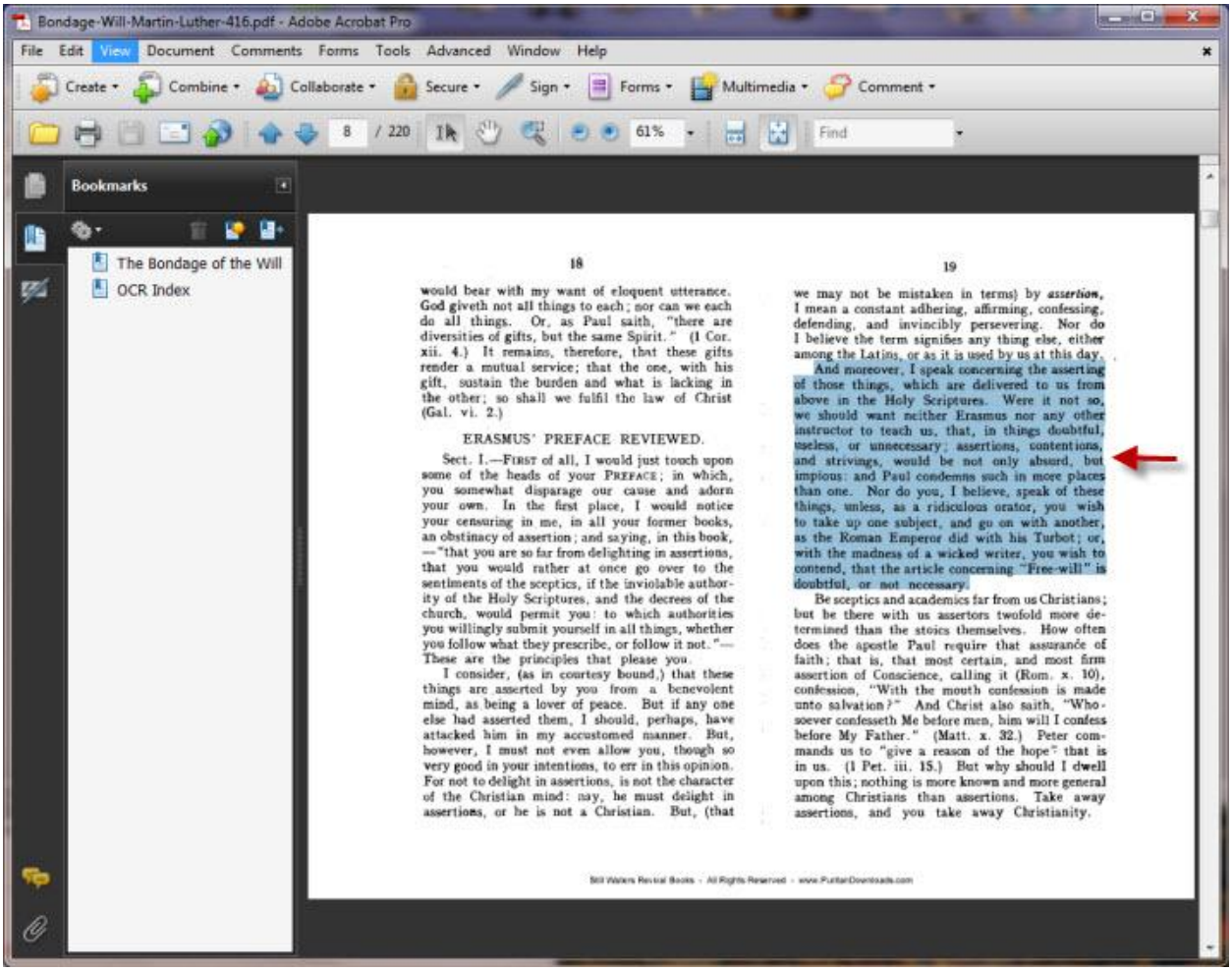


Click on Search Result

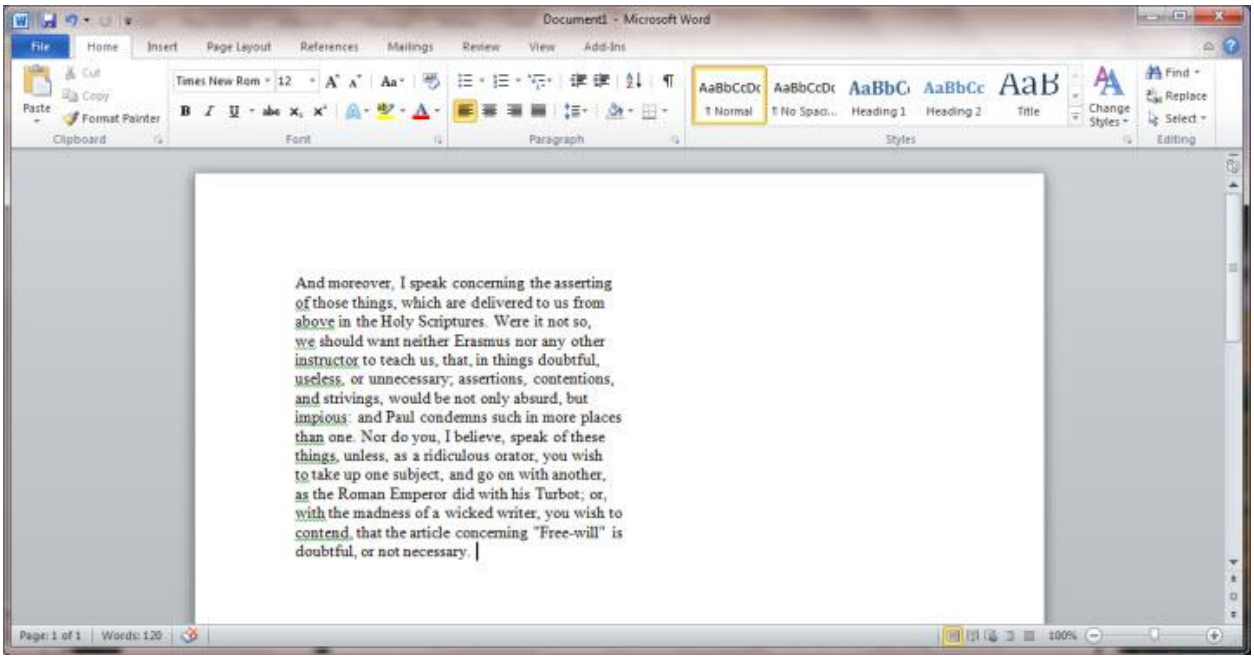


"The Bondage of the Will" Opened to Page Referenced in Search Results

(12) Because the original scanned page images have been processed using Optical Character Recognition (OCR) software, the text that appears in the reader can be copied and pasted into your favorite word processor for editing and inclusion into other documents.

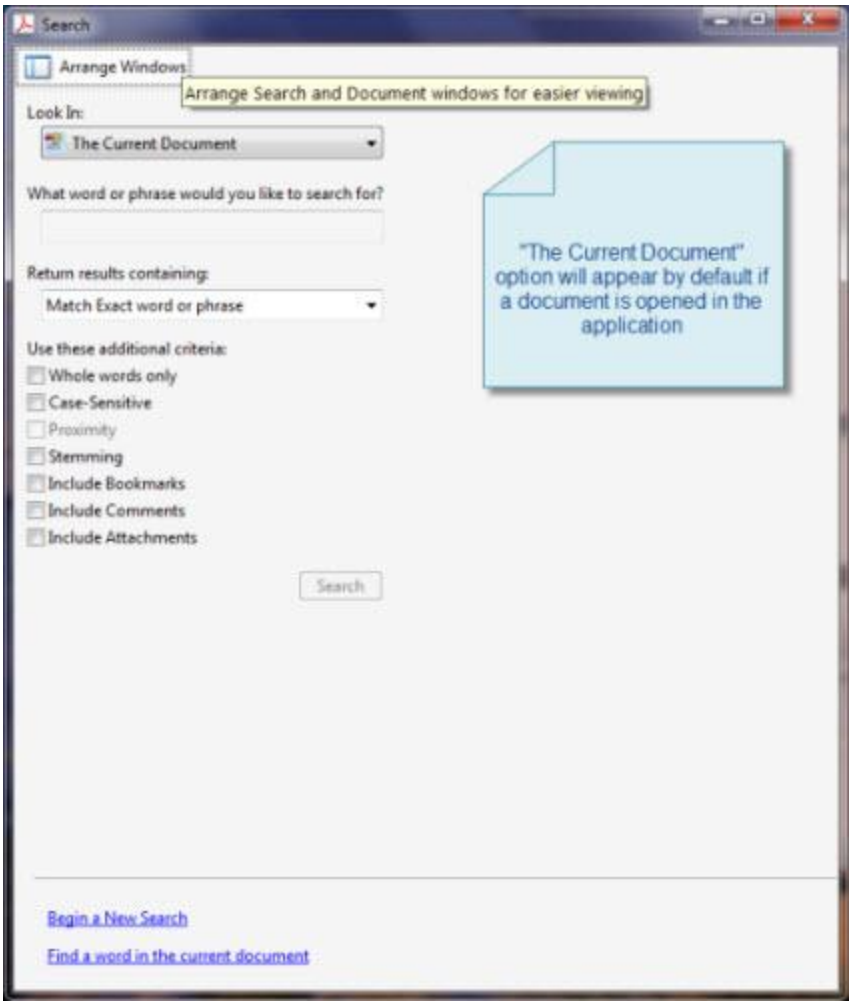


Highlighted Text for Copy & Pasting into Word Processing Software

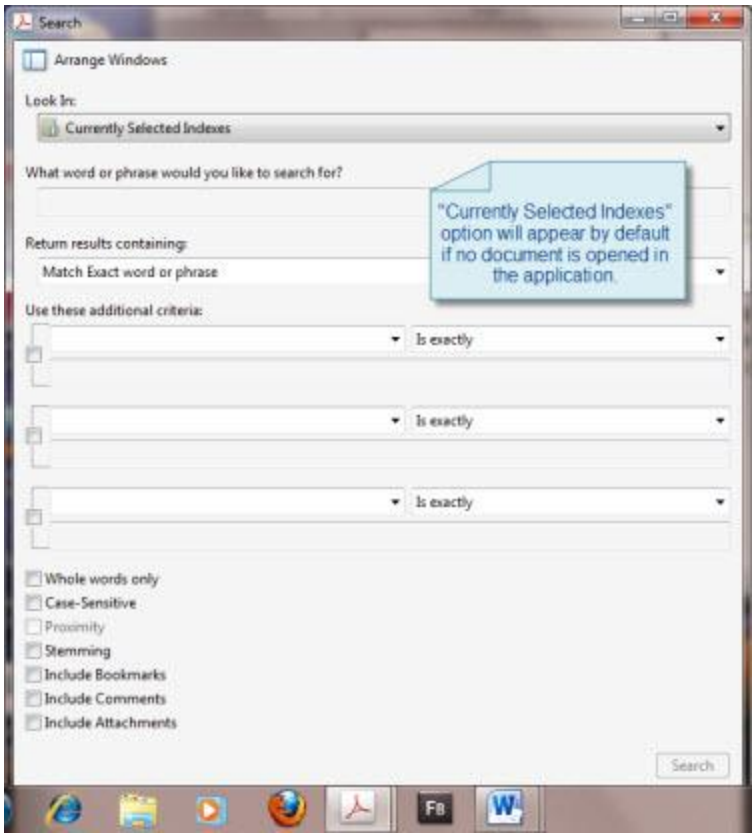


Highlighted Text from "The Bondage of the Will" Pasted Into Microsoft Word for Editing and Inclusion in Another Document

(13) Now that you have set up the Adobe PDF application to use the Master Full Text Index, you need only select the "Currently Selected Indexes" option in the "Look In:" dropdown box to query all the searchable PDFs. If you are interested in searching just the current document opened in the application, select "The Current Document" option in the same dropdown box. Which option appears in the dropdown box by default will depend on the current state of the application. If there is no document opened, you will be presented with the "Currently Selected Indexes" option by default. Conversely, if a document is already opened in the application, the default option will be "The Current Document".



"The Current Document" Default Option



"Currently Selected Indexes" Default Option

(14) Using the above-detailed Master Full Text Index, combined with the extensive searching capabilities of the PHD KnowledgeBase application, the Puritan Hard Drive user is provided with an extraordinarily powerful research/study tool to explore the wealth of resources included in the PHD collection.