



The Plagiarism Resource Site

Charlottesville, Virginia

www.plagiarism.phys.virginia.edu

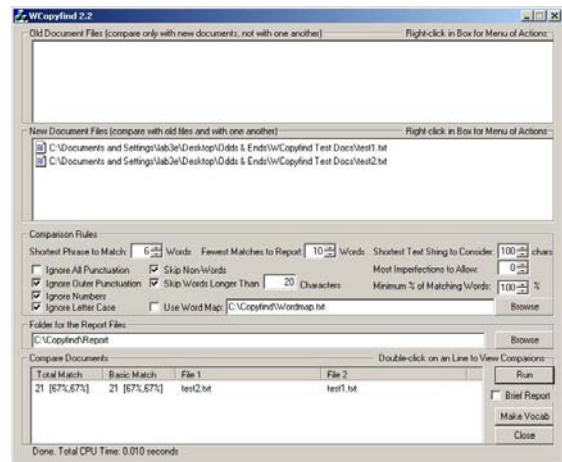
WCopyfind_2.7 Instructions

Purpose:

WCopyfind_2.7 compares text or word processor documents with one another to determine if they share words in phrases.

Overview:

1. Select the documents by browsing for them or dragging them from Windows Explorer into WCopyfind.
2. Adjust the scanning parameters.
3. Select or create a folder in which all the report files will be placed.
4. Run the comparison process.
5. Examine the results.



Step-by-step Instructions:

Step 1: Start WCopyfind_2.7.

- A. Locate WCopyfind and click on its icon

Step 2: Choose Documents to Compare

Two possible paths:

- A. Right-click on the old or new document list.
- B. Select "Browse for Documents" in the popup menu.
- C. Search for and select the documents you want to include in the comparison.
- D. Press "Open" to add these documents to the document list.
- E. Repeat A through D as needed

[Note: Steps A & B can be completed by simply double-clicking on the list.]

[Limitation: Can't handle more than about 500 files at once.]

Or:

- A. Start Windows Explorer.
- B. In Explorer, select one or more documents that you wish to compare. See [Notes](#) for discussion of Web-resident documents.
- C. Drag-and-drop those documents into WCopyfind's document windows.
- D. Repeat B & C as needed.

[Additional Features: You can save or load a particular list of documents via the popup menu (right-click on the list). You can delete selected documents or all of the documents in a list via the popup menu, or selected documents by pressing "delete". Automatic sorting of the documents in a list as they are added can be turned off from the popup menu.]

Step 3: Adjust Comparison Rule Parameters

- A. Shortest Phrase to Match — Range: 0 to infinite

This number is the minimum string length that WCopyfind 2.7 will consider to be a match. For example, when this parameter is set to 6, WCopyfind 2.7 will ignore matching phrases that are only 5 words long or less. I recommend leaving this parameter at 6 (words).

- B. Fewest Matches to Report — Range: 0 to infinite

This number is the fewest matching words in a pair of documents that will cause WCopyfind 2.7 to report a document match in its "Compare Documents" window and generate a pair of underlined comparison documents in the Report Files Folder. There is no recommended value for this parameter.

C. Shortest Text String to Consider — Range: 0 to 255

This number is the shortest sequence of printable characters that WCopyfind 2.7 will extract from a word processor or other document to use in its comparisons. Decreasing it will allow WCopyfind to extract shorter snippets of text, but may cause WCopyfind to include some non-text portions of word processor documents in the comparisons. I recommend leaving this parameter at 100 (characters).

D. Most Imperfections to Allow — Range: 0 to 9

This number is the maximum number of non-matches that WCopyfind 2.7 will allow between perfectly matching portions of a phrase. For example, if this value is set to 2, then WCopyfind 2.7 will bridge its way across two non-matching words to connect pieces of perfectly matching prose. A value of 0 will limit WCopyfind 2.7 to finding only perfect matches, while a value of 1 to 9 will allow WCopyfind 2.7 to find imperfectly matching phrases (matches that contain flaws). Increasing this value slows the program down. I recommend a value of 0 (if speed or absolute matching are your main requirements) or 2 (if you want to find matches despite minor editing).

E. Minimum % of Matching Words — Range: 0 to 100

This number is the minimum percentage of perfect matches that a phrase can contain and be considered a match. Setting this value at 100 limits WCopyfind 2.7 to finding only perfect matches. I recommend a value of 100 (if speed or absolute matching are your main requirements) or 80 (if you want to find matches despite minor editing).

F. Ignore All Punctuation — Checked: Yes or No

When checked, this parameter causes WCopyfind 2.7 to ignore all punctuation characters when it is performing its comparisons. While punctuation will continue to appear in the reports that WCopyfind 2.7 generates, it will not affect the phrase matching. The matching will normally increase when punctuation is ignored. I recommend against checking this box unless you really want to ignore all punctuation.

G. Ignore Outer Punctuation — Checked: Yes or No

When checked, this parameter causes WCopyfind 2.7 to ignore any punctuation characters that appear to the left or right of a word when it is performing its comparisons. For example, the quoted sentence: "The box, which I found, is broken." will be treated as though it were simply: The box which I found is broken (with no final period) . While this "outer punctuation" will continue to appear in the reports that WCopyfind 2.7 generates, it will not affect the phrase matching. The

matching will normally increase when outer punctuation is ignored. I recommend against checking this box if your want absolute matching, but for checking this box if you want to find matches despite minor editing.

H. Ignore Numbers — Checked: Yes or No

When checked, this parameter causes WCopyfind 2.7 to ignore any number characters when it is performing its comparisons. For example, the words 8-fold and 10-fold will match if this parameter is checked. While numbers will continue to appear in the reports that WCopyfind 2.7 generates, they will not affect the phrase matching. The matching will normally increase when numbers are ignored. I recommend against checking this box if your want absolute matching, but for checking this box if you want to find matches despite minor editing.

I. Ignore Letter Case — Checked: Yes or No

When checked, this parameter causes WCopyfind 2.7 to ignore capitalization of letters when it is performing its comparisons. For example, the words Whenever and whenever will match if this parameter is checked. While capital letters will continue to appear in the reports that WCopyfind 2.7 generates, they will not affect the phrase matching. The matching will normally increase when capitalization is ignored. I recommend against checking this box if your want absolute matching, but for checking this box if you want to find matches despite minor editing.

J. Skip Non-Words — Checked: Yes or No

When checked, this parameter causes WCopyfind 2.7 to completely skip words that contain any characters other than letters, except for internal hyphens and apostrophes. The non-words will neither be used in matching, nor will they appear in the reports that WCopyfind 2.7 generates. If you check this box, I suggest also checking "Ignore Outer Punctuation," so that words that begin or end with punctuation aren't skipped over (including plural possessives). I recommend against checking this box if you want absolute matching, but for checking this box if the documents you are comparing contain many non-textual items, including filenames, URL, and other word-processor junk.

K. Skip Words Longer than _____ Characters — Checked: Yes or No, with Range: 0 to 255

When checked, this parameter causes WCopyfind 2.7 to completely skip words that are longer than the number of characters you select. The too-long-words will neither be used in matching, nor will they appear in the reports that WCopyfind 2.7 generates. I recommend checking this box and setting the number of characters at 20, unless your documents really do contain words longer than that. This choice will allow WCopyfind 2.7 to skip over many non-textual items, including filenames, URL, image data, and other word-processor junk.

- L. Use Word Map — Checked: Yes or No, with File Name and Browse Button

When checked, this parameter causes WCopyfind 2.7 to load and use a word map (a generalized thesaurus) of your choice. Once the map has been loaded, WCopyfind 2.7 will examine each word it reads to see if there is a substitute in the word map. It will then perform that substitution prior to doing any comparisons for matching phrases. For example, if the word map indicates that the word "excellent" should be replaced by the word "good", then WCopyfind 2.7 will consider beautiful and pretty to be matching words. Any number of word substitutions is allowed. The original, rather than the substituted words, will appear in any reports generated by WCopyfind 2.7. If you check this box, I suggest also checking "Ignore Outer Punctuation" and "Ignore Letter Case" because the word map requires perfect matching—it considers "Excellent" and "excellent" to be different words. Checking this box will slow the loading and hashing of the documents, but not the comparisons themselves. The format of a word map is described below. I recommend against checking this box unless you know what you are doing, have a good word map file prepared, and want to be able to find matches despite the presence of synonyms.

Step 4: Choose Reporting Folder and Style

- A. Browse to locate or create the reporting folder. It must exist before you run the comparison.
- B. Check "Brief Report" box if you want the comparison files to contain *only* the matching phrases (see [Note](#)).

Step 5: Run Comparison and Examine Results

- A. Click "Run" — matches will be reported in the Comparison Window. (See [Note](#) for explanation of comparison line.)
- B. A small window will open while the comparison process is running, allowing you to abort the process before it finishes.
- C. When the process finishes, a browser window will open, allowing you to examine the pairs of matching files. You can click on the files individually for ease of printing or you can click on the "side-by-side" option to display the pair of file together in adjacent panels of new browser window.
- D. When you view the files side-by-side, all the matching phrases are actively linked between the two files. If you click on a matching phrase in the left file panel, the corresponding phrase in the right file will move to the top of the right panel, and vice versa.
- E. You can also double-click on a comparison line in WCopyfind's internal report window to examine the two comparisons (which have been saved as .html files) in your internet browser.

[Additional Features: You can save the report list to a file via the popup menu (right-click on the report list). You can delete selected lines or all of the lines in the list via the popup menu, or selected lines by pressing "delete". You can also reopen the browser window described in part C.]

Notes:

- A. The comparison window and "matches.txt" both list two numbers of matches. The first or "Total Match" is the number of perfectly matching words that have been marked in the pair of documents. The second or "Basic Match" is the number of perfectly matching words in phrases of at least "Shortest Phrase to Match" words. That second value is essentially the value that would have been obtained if no imperfections were allowed in the matching. In fact, if the "Most Imperfections to Allow" parameter is set to zero, "Total Match" and "Basic Match" will be the same.
- B. The format of a word map is simple. It must be a normal text file (*.txt) containing a series of one-line entries. Each one-line entry should begin with the substituted word, followed by any number of words that are equivalent to that first word. All of these words should be separated by white-space (e.g. blank or tab characters). For example, the line "good excellent terrific wonderful fabulous" will cause any appearance of the words "excellent," "terrific," "wonderful," and "fabulous" to be matched as though the document contained the word "good" instead.
- C. The "Make Vocab" button is intended to help people develop word maps, but it may prove useful in other ways. When you press this button, WCopyfind 2.7 generates a long list of all of the words that appear in all of the document files. It produces an output file, which you chose after pressing the "Make Vocab" button, listing all the words and the numbers of times they appear in all of the documents. The final list is roughly in descending order of usage frequency, though it is not truly sorted. Generating this vocabulary may take a long time, particularly as the list of words encountered gets longer during the generation process. The ignore and skip parameters are active when making a vocabulary and should be selected carefully. I recommend checking "Ignore Outer Punctuation," "Ignore Letter Case," "Skip Non-Words," and "Skip Words Longer than 20 Characters."
- D. In the reports, perfect matches are indicated **by red-underlined words** and bridging, but non-matching words are indicated by *green-italicized-underlined words*.
- E. The matching phrases are links. If you click on a matching phrase, you will be taken to the equivalent phrase in the other document of the pair.
- F. **WCopyfind_2.7 can "surf the web"** by following "internet shortcuts." If you want it to load a document from the web during the comparison process, simply create an "internet shortcut" to that web-document and drag the "internet shortcut" into one of the two input windows of WCopyfind_2.7.

For example, if you want to include <http://www.nytimes.com/> in the documents to be searched, first create an "internet shortcut" to that site somewhere on your desktop or in a folder. You can create it by using an internet browser to open <http://www.nytimes.com/> and then dragging the link icon onto your desktop or into a folder.

You can also create "internet shortcuts" **using a search engine such as Google**. To do this, perform a search using that search engine and then drag and drop each interesting link into a folder on your desktop. The folder will then contain "internet shortcuts" to those interesting pages.

Once you have the "internet shortcuts," you can drag them into WCopyfind_2.7's document windows. You'll see the "internet shortcut" name, followed by the extension ".url". When you run the comparison, WCopyfind_2.7 will load the document over the web and compare it as though it were a local file. You can save a folder full of "internet shortcuts" and drag them into WCopyfind_2.7 whenever you want to compare local documents against them. **Be aware that broken links will terminate the loading process (I should fix this soon)**. Also, WCopyfind reloads web pages when preparing reports that contain them, so if a page changes between the load for comparison and the load for reporting (a very unlikely event, except with news services, etc.), the report may be scrambled.

- G. WCopyfind_2.7 "understands" html-formatted documents, both local and web-resident. If you include a local document that has a .htm or .html extension, or a web-resident document that reports itself as html-formatted, WCopyfind_2.7 will do its best to extract only the text from that file. It should handle html tags almost perfectly, but it is less sophisticated with special characters (it doesn't really understand "&char;"-type instructions).
- H. If you select "Brief Report," WCopyfind_2.7 will abbreviate its report files so that they contain only the matching phrases, in the order in which they appear in the document. Each of these phrases will be followed by a line break. In effect, the "Brief Report" option simply suppresses the inclusion of non-matching text in the reports and also inserts a line break (actually a new paragraph) at the end of each matching phrase.
- I. Each comparison line has four major parts: 1. Total matches, which include too-short phrases that require bridging over non-matching words in order to count as matching between the two documents, 2. Basic matches, which are only those phrases that are long enough on their own to count as matching between the two documents, 3. The first document in the matched pair, 4. The second document in the matched pair. Both match entries (1 & 2) have 3 subparts: a. the number of matching words, b. what percentage of the first document is accounted for by these matching words, ci. what percentage of the second document is accounted for by these matching words.
- J. Starting with this version (2.7), the program tries to handle extremely long file names gracefully. If you still encounter trouble creating or viewing some of the comparison files, consider whether the names of your input files have very long names and whether the report folder has a very long name. Windows cannot handle file names with total lengths more than 255 characters, so it is relatively easy to exceed that maximum. If you are in that situation, try shortening the individual file names or use a report folder with a shorter name.