DoubleKiller Help

1.) Introduction
DoubleKiller's purpose is to find and remove duplicate files. This help file assists you in defining the optimal comparison criteria for your case, selecting folders, setting up appropriate filters for files and, finally, choosing and removing the duplicates. Translations of this help file are available for download at http://bigbangenterprises.de/en/doublekiller.

2.) Comparison criteria
The comparison criteria control how DoubleKiller determines if two files are identical. You can find them on the "Options" page, titled "assume files to be the same, if...". The following criteria are available and can be freely combined:
- "names are identical":
  Compares the file names. This is done case-insensitively, e.g. 'rabbit.ogg' equals to 'Rabbit.OGG'.
- "sizes are identical":
  Compares the file size / length.
- "dates are identical":
  Compares the dates at which the files have been modified the last time. If you create a copy of a file, the new copy gets a new creation date, but the modification date remains the same. However, if you download a file twice, both dates will change.
- "CRC32-checksums are identical":
  Calculates and compares the CRC32 checksums of the files. CRC means Cyclic Redundancy Check. To calculate it, the complete file's content is read and results in a number (the checksum) which can be compared against other checksums.
  You may think of this checksum as a fingerprint. Every file has a fingerprint, and these fingerprints can be easily compared against each other.
  (It is worth to be noted that there is a very small probability of two files having the same checksum although they are not identical, but this is mainly something that can be achieved by strategically manipulating the files to outwit the algorithm which will result in unreadable files. In general you can say that the probability of finding two equally sized, but different files with the same checksum is about as low as finding two humans with identical fingerprints. DoubleKiller Pro supports a byte-per-byte comparison, if you are concerned, though.)

If more than one criterion is selected, the files have to pass all of the selected comparisons to be considered duplicates. The relatively slow process of a CRC32 is always done at last, so in general you should try to take advantage of anything you know about the duplicates you

are searching for by using this knowledge to reduce the amount of CRC32 calculations needed, i.e. minimize the number of files by choosing the right folders and filters (see below) and select as many comparison criteria as possible in your case.
Some hints on which combination to choose for which purpose:

- If you know you have 1:1-copies of files or even complete folders spread on your computer, you should select name, size and date comparison. This will find files you have copied somewhere e.g. for backup purposes without renaming or modifying them in any way. You can also select a CRC32 comparison for additional safety, but this is mostly unnecessary.
- If the above applies, but you may have renamed the files, just deselect name comparison to find identical files with different names. CRC32 comparison is recommended in this case to make sure no false positives are detected which have the same size and date just by coincidence.
- If you have different versions of the same files stored in different folders, but not renamed, you should select only name comparison. In this case it is strongly recommended to only scan the specific folders these files are located in (not your complete hard disk) and to manually check the scan results as there might be files coincidentally sharing the same name.
DoubleKiller Pro offers more features for this case, e.g. a fuzzy filesize comparison to only consider files that are roughly the same size, an advanced filename comparison which can be useful if your file names follow a certain scheme and the possibility of automatically removing e.g. the oldest or smallest files of a dupe set.
- To just scan for any files having exactly the same content without regarding file names or dates, select size and CRC32. This allows to find any intentional or unintentional duplicates, e.g. in your music archive, image collection or downloaded files folder. If you plan to scan your complete disk, it is strongly recommended that you manually check the results and do not touch any files in the Windows folder or other files you are unsure about!

3.) Choosing files / folders to scan

3.1) Selecting Folders

On the "Options" page you can find a box titled "folders". Click the [+] button below it to choose a folder you want to add to the list. You can add as many directories or partitions as you want to and even shared folders of computers of your Network Neighborhood are supported. The checkbox in front of each entry controls if not only files residing directly in this folder but also the files of all its subdirectories are scanned. Click the [-] button to remove the currently selected folder.
If some folders conflict with each other (e.g. if you add 'E:\', including subdirectories, and then 'E:\Music') you will be warned while adding the folder. You'll then have to solve the conflict, otherwise you won't be able to start the scan.
Please do not scan the Windows folder for duplicates (by default called 'C:\Windows', 'C:\WINNT' or similarly) unless you know what you are doing!

3.2) Using file masks

If you are only interested in files of a certain type (more general: files following a certain naming scheme), you can speed up the scan by limiting it to only these particular files. You can also exclude certain files from the scan to protect them.
Mostly you will want to filter the files being scanned based on their file type. In Windows, file names usually consist of a base name and, although it might be hidden by Windows, a short extension (mostly 3 or 4 characters). These two parts are separated by a dot ('.'). Let's take 'Something.bmp' as an example: Its base name is 'Something' and its extension is 'bmp', in this case indicating a bitmap graphic file. To filter for files with a certain extension, you can use file name masks. A mask describes what a file name must look like, using placeholders

for parts of the filename that do not matter. The mask to describe bitmap files is '*.bmp'. The '*' stands for any number of any characters, so '*.bmp' describes any files with the extension '.bmp' and no particular base name.

The masks can be configured in the middle of the "Options" page where you can find two lists of file masks. You can add masks to the lists by entering them below and pressing return. The [-] button deletes the currently selected entry. To edit an entry, doubleclick it. All entries have a checkbox next to them which controls if they are used or ignored. If a list is empty or none of the entries are checked, the list is completely ignored.
To limit the scan to files matching certain masks, add them to the upper list. To have only AVI and MPEG-Files scanned, you would add '*.avi' and '*.mp*g'. This would scan files with the extensions 'avi', 'mpeg' and 'mpg'. If you uncheck or delete all list entries, all files are scanned.
With the second list you can exclude files from the search. You could add '*.dll', '*.sys', '*.vxd', '*.inf' and '*.ini' to this list to protect the most common system files from being deleted. The problem with DLLs (Dynamic Link Libraries) is that many programs need them in their own directory and nowhere else, so by deleting the duplicate DLLs, you might render some of your software unusable.

All masks have to match the following syntax:
- A normal character ('a'-'z' and '0'-'9') in the mask stands for the same character in the file name.
- '?' represents any character, and
- '*' stands for any number of any characters.
- With '[' you can start a set of characters that must be closed by a ']'.
  This set may contain single characters and ranges of characters. One set always represents one character.
  Example: '[A-F0-9ZW]' stands for one character from 'A' to 'F', '0' to '9', 'Z' or 'W'. '[ABCDEF0123456789ZW]' would mean the same.
- A '!' in front of a set means *not*: '[!A-F]' stands for any character except 'A', 'B', 'C', 'D', 'E' and 'F'.
  Masks are always case insensitive.

  A more complex example: '[A-C]??-[0-9][0-9]_?*.txt' describes all files beginning with 'a', 'b' or 'c', followed by exactly two characters (no particular characters, though), a hyphen '-', a two-digit number, an underscore '_' and at least one additional character in front of the extension '.txt'. 'afg-01_Hitchhiker3.txt' and 'Bar-42_F.txt' would fit this pattern, 'boo-2a_new.txt' would not (contains '2a' instead of a two-digit number).

3.3) Filtering by attributes
Since the times of DOS, files and folders can be assigned at least four attributes on Microsoft's operating systems. These are ReadOnly, Hidden, Archive and System. The System attribute is of particular interest as it is assigned to files and folders you should not modify or even delete. Sometimes also the Hidden attribute is used to mark such files.
On the "Options" page, in the box titled "leave out the following files:", you can select any combination of these attributes. If a file or a directory has at least one of the selected attributes assigned, it is excluded from the scan. It is recommended to select the System attribute to protect important system files from being deleted.
In DoubleKiller Pro you can also limit the scan to files with certain attributes instead of excluding them.

3.4) Filtering by file size
In the box titled "leave out the following files:" on the "Options" page you can also exclude files less than (<), equal to (=) or greater than (>) a certain file size. If you only want to scan files having a size e.g. over 1024 kB because small files don't interest you, select "files <

1024 kB".
In DoubleKiller Pro you can define a complete file size range to exclude from the scan or to limit the scan to.

4.) General settings
There are two general settings on the "Options" page in the bottom right corner which have not been described yet.
If you check the checkbox "less detailed status", no detailed status information is shown during the scan, i.e. no file names are displayed. That means you can see the current progress in percent, but messages like "calculating CRC32-Checksum of file: ..." are suppressed. This improves DoubleKiller's performance a little bit.
To save the settings in Windows' registry, select "save settings in registry". If you do not activate this, the default settings are loaded when you restart DoubleKiller and settings like the selected folders are lost. The button below this option removes DoubleKiller's registry entries.

5.) Scanning
Simply click the [Run] button at the bottom of DoubleKiller's window to start the scan.
The scan executes in two steps: First, all files in the selected directories are put into a list, if they are not excluded from the scan due to the criteria you defined. No information on the progress can be displayed as DoubleKiller does not know how many files and subdirectories it will find. The next step is to find out which files are duplicates. Each file is compared with the other files depending on the criteria you selected. The CRC32 is only done if it's unavoidable, to make the program work as fast as possible. Probably you can see the result list filling with files during the scan.
You can click the [Cancel] button or press ESC at any time during the scan to abort it. After asking for confirmation, DoubleKiller usually stops immediately, but you may have to wait a moment if DoubleKiller is just performing a lengthy operation. The result list is not cleared when you cancel the scan and you can work with the list of duplicates discovered so far.

6.) Selecting files to remove
The result list on the "DoubleKiller" page displays all duplicate files found during the scan. The file icons and background colors change between dark and bright every other set of identical files so you can easily distinguish which files belong to a duplicate set. You can doubleclick the entries to open, play or execute the corresponding file. By right-clicking an entry its popup menu is displayed, which, among other options, allows to open the corresponding directory. The status label always displays how many files are currently checked and what their total file size is. There are different ways to select the files to be removed:
One way is to do it completely manually. This takes longer, but gives you the best possible control of which files will be deleted or moved. You can check single files by clicking the icons to the left of the entries. A cross will appear in the icon to indicate that the file is marked for removal. You can also use the keyboard, moving through the list with the up and down buttons and toggling the checkmark with the space bar. You can even combine both input devices by selecting multiple entries with the mouse and using the space bar to check or uncheck them all.
Another way to mark files for removal is to use the buttons to the left of the result list:
- [Check all files]   checks all files in the list
- [Uncheck all files]   unchecks all files in the list
- [Check selected files]   checks all files currently selected (by clicking the entry, not its icon)

- [Uncheck selected files] unchecks all files currently selected
- [Check the first dupes] checks the first file of a pair, the first two of a triple etc.
- [Check the last dupes] checks the last file of a pair, the last two of a triple etc.

The most useful buttons are probably the "check first" and "check last" buttons. With either of these buttons all but one file of every set is marked for removal. You have no great influence on which file of the duplicate sets this will be, though, except that the duplicate sets always follow the same order of paths and therefore all files not removed will mostly remain in the same directories.

Results can be sorted by any criterion which was used to determine if files are identical. So if you were scanning using the default options, you can sort the results by clicking the "Size" or "CRC32" column header. Clicking the same field twice reverts sort order.

DoubleKiller Pro gives you more flexibility here, e.g. you can sort the results pair-by-pair by any criterion. Combined with the "check first" and "check last" buttons you can use this to automatically check the oldest or smallest files of the duplicate sets. Another feature is regrouping of the result list. You can e.g. regroup the list by path to get sets of files in the same directory rather than sets of duplicates, which can be very useful for tasks like deleting files in a certain folder. Additionally, there is a button "check by criteria" available which allows to automatically check/uncheck files depending on the file's sizes, names, folders or dates.

7.) Deleting or moving files

If you checked the files you want to get rid of, you have two options: Either you delete them or you just move them to a backup folder. It is recommended to move the files first and only delete them if you are sure neither you nor any installed software will need any of the files again.

If you decided to delete them click [Delete checked files]. A dialog box appears which displays some information on the checked files and asks if you really want to continue. If you realized that you cannot restore the files once they are deleted (except with some restoration tools, if you are lucky) and still want to continue, click [Yes] and watch the files going into the big bit bucket somewhere up there in heaven.

If you instead want to move the files click [Move checked files...]. After confirming the dialog, you can choose a folder where all checked files will be moved to. It is also possible to create a new folder from within this dialog by entering a nonexistant folder into the edit box or clicking the button on the bottom left of the window (this may not be available on your system). The files' directory structure will be preserved completely. An example: If you choose 'D:\backup' as the destination folder, a file named 'F:\foo\somefile.dat' would be moved to 'D:\backup\F\foo\somefile.dat'. Having selected a folder, you can watch the files being moved to their new home.

DoubleKiller Pro also allows to move the files into Windows' Recycle Bin from which they can easily be restored to their original location or to replace the removed duplicates with shortcuts pointing to the remaining files.

8.) Exporting result list

You can export the complete result list or parts of it into a tab-delimited text file which can not only be read by humans but also imported into your favourite spreadsheet application like OpenOffice.org Calc or Microsoft Excel. The lines are formatted as follows:

    10   Some file.dat G:\Foo\Bar\   21730 2001-02-14 15:48:26 DC0F328B

The first number can be 1 or 0 and indicates if the entry had a bright or dark icon by which you can distinguish the duplicate sets. The second number can be 1 or 0, too, and indicates if the entry was checked or unchecked. Delimited by tabulators the file name, path, size in bytes, date and time in your local format and the CRC32 checksum follow (the latter is only included if a CRC32 comparison has been performed).

To export the complete result list, right-click any of the entries and choose "Export all

entries...". You will be presented a dialog asking you where to save the file.
To only export a part of the list, select the entries you want to export first. Multiple entries can be selected by holding the CTRL key and clicking them, a range of entries can be selected by clicking the first one in the range, holding the SHIFT key and clicking the last one. When you have selected all entries you are interested in, right-click any of the selected entries and choose "Export selected entries...". You can also choose "Copy selected entries" to copy them into the clipboard rather than saving them into a file.

It is not possible to import result lists into DoubleKiller, but in DoubleKiller Pro it is possible to export the result list into a binary format which can be reloaded at any time, enabling you to continue or review your work later.

9.) Uninstall
As DoubleKiller does not need to be installed, only extracted and run, there is no uninstall entry in Window's control panel. To remove it, simply perform the following steps:
- Start DoubleKiller, switch to the "Options" page.
- At the bottom right there is a checkbox "save settings in registry". If this checkbox is checked, click the button [delete settings from registry]. If it isn't checked, DoubleKiller never wrote anything to the registry so nothing has to be tidied up.
- Close DoubleKiller.
- Delete "DoubleKiller.exe" and you're done.

10.) DoubleKiller Pro
DoubleKiller Pro is a heavily extended version of DoubleKiller featuring superior performance, higher flexibility and new features. Its additional features include:
- Multilingual interface: With freely downloadable language files the complete interface can be translated into another language
- Advanced comparison options for each of the four basic comparison criteria, e.g. to compare only a particular part of the file name or content, to allow a specified tolerance for file size or date, or to perform a byte-per-byte comparison instead of a CRC32
- Improved folder selection: Directories can be excluded from the scan (e.g. to scan 'C:\', but exclude 'C:\Windows') or temporarily disabled (no need to remove them from the list)
- A distinction between so-called *fresh* and *library* folders allows to quickly compare a small number of unknown files against a huge collection
- The result list can be sorted pair-by-pair or regrouped by any criterion - combined with the "check first" and "check last" buttons this allows to automatically mark e.g. the oldest or smallest files of a set for removal
- There is a new button "check by criteria" to check or uncheck files based on criteria like file name masks, path masks, file size or date ranges
- Beside moving the checked files into a specified folder or directly deleting them, you can move them into Windows' Recycle Bin from which they can easily be restored to their original location or replace the duplicates by shortcuts pointing to the remaining files
- Directories that have been emptied completely while removing duplicate files can be deleted automatically
- You can save the result list at any time and continue or review your work later
- Command line support: The complete process of selecting folders, scanning for duplicates and removing files can be automated

DoubleKiller Pro can be purchased for $19.95 / 14.95 € at http://bigbangenterprises.de/en/doublekillerpro. There is a free demo version available so you can check out the new features.
So if you like DoubleKiller and want to support it, try DoubleKiller Pro!

11.) Contact

Please contact me at mailto:info@bigbangenterprises.de if you have any suggestions, found bugs or just want to tell me your opinion about DoubleKiller. Every kind of feedback is appreciated and will be answered!
Updates for DoubleKiller can be found at http://bigbangenterprises.de/en/doublekiller. Please subscribe to the newsletter at http://bigbangenterprises.de/en/newsletter.htm if you wish to be notified when a new version of DoubleKiller is released.

12.) Disclaimer

DoubleKiller is freeware for private use and you may freely use, copy, and distribute this software, provided that neither a fee is charged nor donations are requested for such use, copying or distribution, it is not used commercially or executed on a computer that is used commercially and it is distributed in whole and in its original, unmodified state.
DoubleKiller can be used to permanently delete files from your computer. Although it is thoroughly tested and will never delete files you did not mark for removal, you can damage your system or lose data when you delete files. To the extent allowed by applicable law, DoubleKiller is therefore provided "as is" without warranty of any kind, whether express or implied, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. In no event shall the author be liable for any direct, indirect, consequential, incidental, special, punitive or other damages whatsoever, including, but not limited to, loss of business profits or information, business interruption or other pecuniary loss, even if the author has been advised of the possibility of such damages.
Designated trademarks and brands in this documentation and any other parts of the software are the property of their respective owners.