

DuplicateFileDetective™

Professional Grade Duplicate File Management



Duplicate File Detective Help

© 2003-2009 Key Metric Software, All Rights Reserved

Table of Contents

Foreword	0
Part I Introduction	4
1 Introducing Duplicate File Detective	4
2 Getting Started.....	4
3 How to Buy.....	4
4 Product Registration.....	5
Part II Duplicate File Detective Projects	5
1 What is a Project?.....	5
2 Managing Projects.....	6
3 Project Wizard.....	6
4 Search Paths.....	7
5 Exclusions.....	8
6 Search Filtering.....	9
About Search Filtering	9
Filename Matching	9
File Dates	11
File Sizes	12
Advanced Options	12
7 More Project Settings.....	12
Match Settings	12
Part III Duplicate Results	14
1 Duplicate Search Progress.....	14
2 Result Report.....	14
3 Groups.....	15
4 Sorting.....	15
5 File Checksums.....	15
6 Context Menu	16
7 Marking Files.....	17
8 SmartMark.....	17
9 Moving, Zipping & Deleting.....	18
10 Exporting.....	19
11 Importing.....	20
12 Summary Report.....	21
13 Autoprune.....	21
14 Printing.....	21
15 Result Filtering	22

16	Image Preview	22
17	Folder View	22
18	File Properties.....	24
Part IV File Types		25
1	About File Types.....	25
2	File Types Report.....	25
3	File Types Graphs.....	26
4	File Type Researcher.....	26
Part V Preferences		27
1	About Preferences.....	27
2	General.....	27
3	Protection.....	27
4	Results Report.....	28
5	Update Checking	29
6	New Projects.....	29
7	File Types.....	29
8	Logging	30
9	Advanced.....	30
Part VI Advanced Features		31
1	Customizing.....	31
2	Command Line Options.....	32
3	File Hash Calculator	33
4	Error Log.....	33
5	Managing File Type Groups.....	33
Index		34

1 Introduction

1.1 Introducing Duplicate File Detective



Duplicate File Detective is a powerful tool designed to help you find and manage duplicate files on any Windows file system.

Duplicate File Detective will help you to reclaim valuable disk space by providing you with a wealth of tools and reporting options, as described in the remainder of this help file.

As you use Duplicate File Detective, please keep in mind that your feedback is extremely valuable to us. Feel free to email us at any time at feedback@duplicatefiledetective.com. We'd appreciate hearing from you.

1.2 Getting Started

To starting using Duplicate File Detective, click the Project Wizard button located in the Project section of the application Ribbon Bar. You will be asked to choose a File Matching method - just click the Next button for now to use the default Checksum Duplicate Scan.

Next, the Project Wizard will ask you to specify which drives or folders you would like to scan. For now, select "My Documents" from the listing on the right and click the blue right arrow button to move "My Documents" to the list of directories to scan. Click the Next button again.

Project Wizard will now ask you to specify which types of files you would like to scan. For now, just accept the default file matching pattern (*.*) and click the Next button.

Finally, the Project Wizard will tell you that it's ready to start. Click the Finish button, and you're off and running. The remaining sections of this help file will talk about all the things that you can do from this point forward.

Tip: You certainly don't have to use the Project Wizard if you're prefer not to. The remainder of this help file will show you everything you need to know to configure a project manually.

1.3 How to Buy

Duplicate File Detective is a commercial software product with a 15-day trial period. This means that you have 15 days from the time you install Duplicate File Detective to decide whether or not you find it worth purchasing.

For more information on how to purchase Duplicate File Detective, please visit our product website at <http://www.duplicate-file-detective.com/>. A wide variety of purchasing methods are available.

Also please feel free to contact us with any sales questions you might have. Just send an email to

sales@duplicate-file-detective.com, and we'll get back to you promptly.

Thank you for trying Duplicate File Detective.

See also:

[Product Registration](#)

1.4 Product Registration

To continue using Duplicate File Detective beyond the initial 15-day trial period, you must [purchase](#) one or more product licenses.

We offer a variety of licensing options, including single-user, multi-user, site-wide, and enterprise-wide licenses. Please see the "Order" section of our [product website](#) for details.

Once you complete the product license purchase process, you will be provided with a license name and key which you can then enter into Duplicate File Detective to "unlock" it for permanent use. When entering your product license name and key, please keep in mind that they must be entered *exactly* as they were provided to you. The easiest way to ensure correct license entry is to copy and paste these values using the Windows clipboard.

If you have any problems entering your product license details, feel free to contact us at support@duplicate-file-detective.com. We're happy to assist our valued customers.

Upgrade Policy

Upgrades within the same major product version (i.e. from v3.0 to v3.1) are free.

Major version upgrades (i.e. from v3.0 to v4.0) are made available at significant discount to existing product license holders.

See also:

[How to Buy](#)

2 Duplicate File Detective Projects

2.1 What is a Project?

In Duplicate File Detective, a Project is comprised of all the attributes that define the current duplicate file scan. These include the current filter criteria (such as filename masks, etc.), [search paths](#), [exclusion folders](#), [file matching methods](#), and more.

One of Duplicate File Detective's most powerful features is its ability to save and load project as needed. Duplicate File Detective does this with *project files*, which have a filename extension of .dfd (e.g. "my-dup-project.dfd"). Double-clicking a project file in Windows Explorer will cause Duplicate File Detective to load the project.

For more information, see the [Managing Projects](#) section of this help file.

See also:

[Managing Projects](#)
[Project Wizard](#)

2.2 Managing Projects

Duplicate File Detective projects are managed with a range of commands accessible via the application Ribbon Bar. These include:

- **Project Wizard** - Used to quickly create and launch a new Duplicate File Detective project. This is a great way to get started.
- **New Project** - Used to create a new project with default search folders, exclusions (excluded folders), and other project settings. The project can then be modified in whatever way suits your needs.
- **Open Project** - Allows you to open an existing project file (e.g. "filename.dfd").
- **Save Project** - Used to save the active project. The Ribbon Bar button will appear "enabled" only when unsaved changes to the current project are pending.
- **Run Project** - Used to execute the currently defined project.

Tip: You can also load projects by double-clicking on a project filename (a file with a .dfd extension) from outside of Duplicate File Detective (from Windows explorer, for example).

See also:

[What is a Project?](#)
[Project Wizard](#)

2.3 Project Wizard

The Duplicate File Detective Project Wizard is designed to provide you with a quick starting point for duplicate file scans.

Start the Project Wizard by clicking the appropriate button in the application Ribbon Bar.

Step 1 - File Matching

The first screen of the wizard helps you to define the type of scan to execute; there are several types.

- **Quick Duplicate Scan** - scans the selected drive(s) or folder(s) for duplicates that match by file name and size. This provides quick results with a fair degree of accuracy.
- **Checksum Duplicate Scan** - scans the selected drive(s) or folder(s) for duplicates that match by file name, size and 32-bit CRC content checksum. Quick, with a good degree of accuracy.
- **Strong Checksum Duplicate Scan** - scans the selected drive(s) or folder(s) for duplicates that match by file size and 128-bit MD5 content checksum (*without regard to file names*). Slower, with a very strong degree of accuracy.

To select a one of the file matching methods listed above, click the blue hyperlink text associated with

the name.

Step 2 - Scan Paths

The next screen of the Project Wizard allows you to determine which file system paths should be scanned for duplicate files. You will be presented with two panes - a file system browser (on the left) that shows available drives and folders (similar to the presentation in Windows Explorer), and another pane (on the right) that lists directories to be scanned.

To select a drive or folder for scanning, simply select it in the file system browser pane and click the right blue arrow button.

The Directories pane on the right is also a full-fledged path editor; you can use its toolbar buttons to add, edit, or re-order paths that appear in this pane.

Add as many paths as you like during this step in the project wizard.

Step 3 - File Types

The Project Wizard file types screen allows you to specify precisely what kinds of files should be included in the duplicate scan process. For example, if you'd like the duplicate scan to include only certain types of multimedia files, you can do so here.

The file types input field uses Duplicate File Detective's [filename masking system](#), which supports both wildcards and regular expressions.

Step 4 - Ready to Begin

Duplicate File Detective is now ready to begin scanning for duplicates! Click the Finish button, and watch as Duplicate File Detective goes to work!

Note that, depending upon selections made in previous steps (and the size of the file system paths being scanned), the duplicate file identification process can be fairly time consuming. Once you've got the basics down, learn how to use [Search Filtering](#) to limit the files being compared, which can accelerate the process considerably.

See also:

[What is a Project?](#)

[Managing Projects](#)

2.4 Search Paths

Duplicate File Detective provides the ability to scan multiple duplicate file search paths in a single pass. These paths are managed via the Search Paths docking window, located (by default) in the upper left hand corner of the Duplicate File Detective application window.

Managing Search Paths

The toolbar buttons near the top of the search paths docking window pane provide the ability to add,

delete, protect (more on this in a minute) and alter the position of search path entries. Search paths are processed in the order in which they appear in the search path listing.

Note that each search path in the search path listing has a check box to its left. Duplicate File Detective will only process paths that have a mark in this check box. This provides a quick way to disable or enable paths during search operations without removing and re-adding them.

Right-clicking within the Search Paths window will cause a special context-sensitive menu to appear. This menu provides another means of manipulating search paths, and also allows you to check / uncheck all paths at once. Double-clicking a search path will edit that path.

When a new Duplicate File Detective project is created, the search path listing is populated with a list of default values correlating to the fixed drive on the host computer. This behavior can be altered in the [preferences](#) window.

Search paths are saved and loaded on a per-project basis.

Tip: Network paths can be entered into the Search Paths window by clicking the "Add" toolbar button and either selecting a path from the "My Network Places" branch of the file system tree, or by entering a UNC path (e.g. "\\myserver\sharename") directly into the Folder edit box.

Another tip: You can also drag folder paths from within Windows Explorer and "drop" them on the Search Paths panel.

Protecting Search Paths

Search paths can also be protected from processing by selecting the path in question and clicking the Protect toolbar icon.

When a search path is protected, any duplicate file found within that path (or any of its subfolders) will also be protected. Protected entries will be shown in the duplicate [results report](#) view with a small lock icon in place of the normal check box image that appears adjacent the file name. This effectively prevents such entries from being marked, and therefore prevents them from being [processed](#) by Duplicate File Detective.

Note that search paths can be protected either before or after a duplicate file search is run.

Important: Search paths (and therefore search path protection) is a property of the current Duplicate File Detective [project](#). As such, duplicate search results [imported](#) from an existing XML [export](#) file will *never* be identified as protected unless the associated project is also loaded.

See also:

[Exclusions](#)

2.5 Exclusions

Similar in form and function to the [Search Paths](#) docking window pane, the Exclusions pane allows you to indicate which file system paths should be excluded from duplicate file search operations.

Note that, like the search paths, a default set of excluded paths is provided for you when a new project

is created. This behavior can be altered in the [preferences](#) window if desired. Generally speaking, it is recommended that you exclude file system paths that may contain sensitive files such as those used by the Windows operating system or third-party software. This can also dramatically reduce the working file set and produce faster duplicate file search results.

Please review the [Search Paths](#) section of this help file for usage details.

See also:

[Search Paths](#)

2.6 Search Filtering

2.6.1 About Search Filtering

Duplicate File Detective uses Search Filtering to determine which file system objects should be subject to the duplicate comparison process. When a Duplicate File Detective project is run, each file is compared against the current search filter criteria - if the file doesn't match, then it is excluded from the duplicate comparison process.

Search Filtering settings are accessed through the docking panel visible to the right of the Duplicate File Detective user interface (by default). If the panel is not visible, click the View menu and select the Search Filtering menu item.

There are four main search filter criteria types, which are further discussed in other parts of this guide.

- [Filename masks](#) - Compare only the file names and types in which you're interested.
- [File dates](#) - Compare only files of a certain age.
- [File sizes](#) - Compare only files of a certain size.
- [Advanced options](#) - Advanced file scanning / comparison options.

Each of these filter criteria types is contained within its section within the Search Filtering docking panel. Clicking the header text for any section will cause its contents to be expanded or collapsed.

Tip: Whenever possible, use Search Filtering settings to narrow down the number of files that are compared for duplication. Doing so will increase the speed of the duplicate comparison process.

2.6.2 Filename Matching

As duplicate file system scanning progresses, each file name encountered is compared against the current project's filename masks. If the mask does not match the name of the current file, that file is not included in the duplicate comparison process.

Filename masks can be defined using one of two syntaxes: basic wildcard patterns or more advanced regular expressions. Use the "Use regular expression matching" switch in the filename masks section of the [search filtering](#) panel to switch between these two modes.

Wildcard Patterns

When operating in wildcard pattern mode, filename masks are comprised of patterns of characters, with multiple sets of masks separated by a semicolon. Wildcard characters include '?' and '*', which match

either one instance or multiple instances of any character (respectively). Any other (non-wildcard) character matches itself.

Further, any filename mask can be preceded by a tilde character ('~'), which specifies that the mask is exclusionary. If a filename matches an exclusionary mask, the file will always be skipped.

Wildcard Pattern Examples

Here are a few examples of wildcard filename masks:

- **mypicture.bmp** - This mask has no wildcard characters, and is therefore a literal match. Only files named "mypicture.bmp" will be included in the duplicate comparison process.
- ***.bmp** - This mask uses the asterisk (*) character to include any files with a ".bmp" extension in their names.
- ***.bmp;*.gif;*.jpg** - This is a compound mask, with separate entries separated by semicolons. This mask will match any ".bmp", ".gif", or ".jpg" files it encounters.
- **~family*;*.bmp** - The first element of this compound mask is preceded by a tilde (~), which means that it will excluded any files that start with "family". It will then include any files with a ".bmp" extension in their names.

The filename masks section of the Search Filtering docking pane also includes a **Presets** button, which you can click for easy access to a range of built-in masks. These can help you to get started quickly. Presets can also be [customized](#) to suit your needs.

Important Note: When using wildcards in filename masks, keep in mind that a mask of *.* is subtly different than just *. The former requires that a dot (.) be included in the filename, while the latter does not. In other words, the *.* pattern will not match filenames that don't have an extension - if you wish to match these, use just a single asterisk (*) instead.

Regular Expressions

As noted above, filename masks can also be defined using powerful regular expression syntax. Regular expressions are formulas that can be used to match strings of text that follow some pattern. They allow their users to succinctly express a set of character matching rules that would otherwise require a large number of switches and logical operations.

This help file will not provide an in-depth tutorial on the formation of regular expressions, simply because a large number of such tutorials are freely available on the web (visit your favorite search engine and enter "regular expressions" into the search box to find them).

Please keep in mind, however, that there are subtle differences between the regular expression syntax engines that various applications employ. The charts below provide an overview of the regular expression metacharacters and abbreviations supported by Duplicate File Detective.

Metacharacter	Meaning
.	Matches any single character.
[]	Indicates a character class. Matches any character inside the brackets (for example, [abc] matches "a", "b", and "c").
^	If this metacharacter occurs at the start of a character class, it negates the character class. A negated character class matches any character except those inside the brackets (for example, [^abc] matches all characters except "a", "b", and "c"). If ^ is at

	the beginning of the regular expression, it matches the beginning of the input (for example, <code>^[abc]</code> will only match input that begins with "a", "b", or "c").
-	In a character class, indicates a range of characters (for example, <code>[0-9]</code> matches any of the digits "0" through "9").
?	Indicates that the preceding expression is optional: it matches once or not at all (for example, <code>[0-9][0-9]?</code> matches "2" and "12").
+	Indicates that the preceding expression matches one or more times (for example, <code>[0-9]+</code> matches "1", "13", "666", and so on).
*	Indicates that the preceding expression matches zero or more times.
??, +?, *?	Non-greedy versions of ?, +, and *. These match as little as possible, unlike the greedy versions which match as much as possible. Example: given the input " <code><abc><def></code> ", <code><.*?></code> matches " <code><abc></code> " while <code><.*></code> matches " <code><abc><def></code> ".
()	Grouping operator. Example: <code>(\d+,\d+)</code> matches a list of numbers separated by commas (such as "1" or "1,23,456").
\	Escape character: interpret the next character literally (for example, <code>[0-9]+</code> matches one or more digits, but <code>[0-9]\+</code> matches a digit followed by a plus character). Also used for abbreviations (such as <code>\a</code> for any alphanumeric character; see table below). If <code>\</code> is followed by a number <code>n</code> , it matches the <code>n</code> th match group (starting from 0). Example: <code><{.*?}>.*?</0></code> matches " <code><head>Contents</head></code> ".
\$	At the end of a regular expression, this character matches the end of the input. Example: <code>[0-9]\$</code> matches a digit at the end of the input.
	Alternation operator: separates two expressions, exactly one of which matches (for example, <code>T t</code> matches "The" or "the").
!	Negation operator: the expression following <code>!</code> does not match the input. Example: <code>alb</code> matches "a" not followed by "b".

Abbreviations

<code>\a</code>	Any alphanumeric character: <code>([a-zA-Z0-9])</code>
<code>\b</code>	White space (blank): <code>([\t])</code>
<code>\c</code>	Any alphabetic character: <code>([a-zA-Z])</code>
<code>\d</code>	Any decimal digit: <code>([0-9])</code>
<code>\h</code>	Any hexadecimal digit: <code>([0-9a-fA-F])</code>
<code>\n</code>	Newline: <code>(\r \r?\n)</code>
<code>\q</code>	A quoted string: <code>(\"[^\"]*\")(\'[^\']*\')</code>
<code>\w</code>	A simple word: <code>([a-zA-Z]+)</code>
<code>\z</code>	An integer: <code>([0-9]+)</code>

Excluding Protected File Types

By default, Duplicate File Detective will exclude [protected file types](#) from the duplicate search process. To manage protected file types, navigate to the Protection tab of the Preferences window.

2.6.3 File Dates

During the duplicate file search process, every file's date/time stamp is examined and compared against the settings defined in the **File dates** section of the [Search Filtering](#) docking panel.

File date comparison options include:

- **Include all dates** - No file date filtering is performed.

- **Modified within last week** - Include only files with a date stamp that occurs within the last week.
- **Modified within last month** - Include only files with a date stamp occurring within the last month.
- **Modified with last year** - Include only files with a date stamp occurring within the last year.
- **Specify date range / type** - More info below...

If you select the last option (specify date range / type), a drop-down box will be enabled that allows you to choose the file date / time stamp type used during comparison operations. Choices include date modified, date created, and date accessed. You will also be able to define the specific date range within which file date / time stamps must fall in order to be included in the duplicate comparison process.

2.6.4 File Sizes

During the duplicate file search process, every file's size is noted and compared against the settings defined in the **File sizes** section of the [Search Filtering](#) docking panel.

File size comparison options include:

- **Include all sizes** - No file size filtering is performed.
- **Small** - Only small files (those less than 100KB in size) will be included.
- **Medium** - Only medium files (those less than 1MB in size) will be included.
- **Large** - Only large files (those more than 1MB in size) will be included.
- **Specify size** - More info below...

If you select the last option (specify size), you will be able to specify that only files that are "at least" or "at most" a specific size should be included.

2.6.5 Advanced Options

Advanced [Search Filtering](#) options are as follows:

- **Search system files / folders** - Indicates that system files and folders should be included. Off by default.
- **Search hidden files / folders** - Indicates that hidden files and folders should be included. Off by default.
- **Search sub folders** - Indicates that subfolders should be searched recursively. On by default.
- **Case sensitive masks** - Indicates that [filename masks](#) should be treated as case-sensitive. Off by default.
- **Skip zero-length files** - Specifies that zero-length files should be skipped. On by default.

Tip: You should leave the advanced options at their default settings unless you have a good reason to do otherwise.

2.7 More Project Settings

2.7.1 Match Settings

The Match Settings window provides more granular control over how Duplicate File Detective's compares files.

File Comparison Options

- **Match file names** - Uses the filename portion of a given search path.
 - **Ignore file extensions** - Uses the filename portion of the search path, but without regard to the file extension.
- **Match file sizes** - Compares the precise size of files.
 - **Match file contents** - Uses one of the comparison hash types discussed below to represent and compare file contents.
 - **Byte-for-byte content match confirmation** - Confirms that matches identified by content hashing are identical at the byte level.
- **Match last modified date / time** - Uses a combination of date / time to compare files.

Note that byte-for-byte content matching will slow the overall duplicate search process considerably, and is rarely necessary (see file hashing notes below).

File Content Hash Types

The File Match Settings window further allows you to configure the type of *file hash* used for content comparison operations. A file hash is a numerical checksum value, derived through some mathematical formula, that represents the contents of the related file as a whole. Theoretically speaking, stronger file hash algorithms produce checksums that are more unique than weaker ones, and thus are more likely to correctly identify duplicate files. Generally, the stronger the file hashing algorithm, the longer it takes to produce a file checksum.

Note: the *match file sizes* option must be engaged in order to enable the *match file contents* option. The contents of two files cannot be considered identical unless the files are of the same size.

Duplicate File Detective supports the following file comparison hash types:

- **CRC32** - A quick, 32-bit checksum.
- **ADLER32** - Another 32-bit checksum, similar in accuracy to CRC32.
- **MD5** - A very accurate, slower 128-bit checksum.
- **SHA1** - Even more accurate, slower 160-bit checksum.
- **SHA256** - Even more accurate, slower 256-bit checksum.
- **SHA512** - Even more accurate, slower 512-bit checksum.

Note that stronger file content hashing algorithms such as MD4 and SHA1 are very *unlikely* to produce false positives (e.g. mistakenly identify two files as being identical to one another when they actually differ). Even the smallest differences in file contents will (with overwhelming probability) result in completely different hashes due to a cryptographic concept known as the [avalanche effect](#). If you must be absolutely certain that two files are identical, use the byte-for-byte content match confirmation, which validates file comparisons at the binary level.

File Matching options are project-specific, and are saved and loaded on a per-project basis.

See also:

[File Checksums](#)

3 Duplicate Results

3.1 Duplicate Search Progress

When a Duplicate File Detective project is run, a progress window will appear for the duration of the duplicate file scan and comparison process.

The progress window shows a variety of information, including the number of folders scanned, folders skipped, files scanned, files compared, and duplicates found. It will also show you the name of the current folder being scanned, or a pair of progress bars that indicate the state of the overall file comparison process.

Duplicate file scanning and comparison operations can be stopped at any time. If a scan is aborted before it completes, the results gathered so far will be displayed.

Note: Folders and / or files are generally "skipped" because of the project [Search Filtering](#) settings currently in effect. If an entire folder is skipped (because it is [excluded](#), for example), then the files contained within the folder will not be counted. For details on what file and folders are skipped, refer to [Logging](#).

3.2 Result Report

The Duplicate File Detective result report provides a group-based listing of duplicate files found on the scanned file system(s).

The result report has a number of powerful features, including:

- [Exporting](#) - Export the duplicate file report in a range of formats, include XML (which can be [re-imported](#)).
- [Printing](#) - Printing, print preview, and printer setup.
- [Groups](#) - Duplicate files are separated into groups to facilitate comparison and management.
- [Sorting](#) - Sort duplicate files (and groups) by any search result report column.
- [Context menu](#) - Right-click to access numerous file and group level functions.
- [Customization capabilities](#) - Used to make the duplicate report work the way you want.
- [File marking](#) - Used to mark files for further processing.
- [Moving, Zipping & Deleting](#) - Built-in duplicate file processing assistance.

Click one of the links above to further explore the capabilities of the Duplicate File Detective result report window.

Note: Many aspects of the duplicate results report display can be customized. See the [Customization](#) section of this help file for additional information.

See also:

[Duplicate Groups](#)
[Sorting](#)
[Context Menu](#)
[Marking Files](#)
[Result Filtering](#)

3.3 Groups

Once duplicate file searching and comparisons are complete, the [result report](#) will provide a group-based display of the results.

Duplicate file groups are separated by alternating background colors. You can also hover your mouse over any file within a duplicate group to get information about the overall size of that group.

Groups also have a unique number (a group index) associated with them, visible (by default) as a column within the result report.

Which Duplicate File Group Entry is the Original?

When viewing and managing duplicate file groups, don't be confused by the concept of duplicate file *originality*. Duplicate File Detective has no way of knowing which duplicate file in any given group is the "original", and it doesn't attempt to make such determinations - it simply shows all files discovered during the duplicate search process. Decisions regarding which files to keep (or otherwise [process](#)) are left entirely up to the user.

See also:

[Marking Files](#)

3.4 Sorting

You can sort the duplicate [result report](#) by clicking any of its column headers. Click once to sort by the column in ascending order, again to sort in descending order.

How Sorting Works

Duplicate File Detective implements a highly specialized approach to sorting duplicate search results. Specifically, sorting is done at the [duplicate file group](#) level in a manner that ensures that duplicate groups are kept together within the results report.

Say, for example, that you sort the current duplicate search results by size in descending order (so that the largest files are near the top of the report). First, the files within each duplicate group will be sorted (by size in descending order), and then the duplicate file groups themselves will be sorted in the same manner, using the size of the first duplicate file within the group (which will be the largest due to the previous step) as the sort key.

If that seems confusing, please spend some time playing with sorting duplicate search results to see how this works. The key concept to remember is that Duplicate File Detective will **always keep duplicate file group entries together** during sorting operations.

3.5 File Checksums

Duplicate File Detective can compare the contents of files by computing the file's content **checksum**. This checksum is a numerical representation of the file's contents derived through a series of mathematical computations - a process known as *hashing*.

Duplicate File Detective offers a range of hashing algorithms:

- **CRC32** - Quick checksum (32 bits)
- **ADLER32** - Quick checksum (32 bits)
- **MD5** - Stronger, slower checksum (128 bits)
- **SHA1** - Even stronger, slower checksum (160 bits)
- **SHA256** - Even stronger, slower checksum (256 bits)
- **SHA512** - Even stronger, slower checksum (512 bits)

Generally speaking, the "stronger" the checksum / hashing method, the more likely it is that two files compared with the resulting checksum will be identical. Stronger hashing algorithms are also generally a bit slower than weaker ones.

Note that stronger file content hashing algorithms such as MD5 and SHA are *extremely unlikely* to produce false positives (e.g. mistakenly identify two files as being identical to one another when they actually differ). Even the smallest differences in file contents will (with overwhelming probability) result in completely different hashes due to a cryptographic concept known as the [avalanche effect](#). If you must be absolutely certain that two files are identical, use the byte-for-byte content match confirmation, which validates file comparisons at the binary level.

Tip: Duplicate File Detective provides a [File Hash Calculator](#) feature that you can use to experiment with the computation of file checksums.

See also:

[File Matching](#)

3.6 Context Menu

Right-clicking within the Duplicate File Detective [result report](#) will cause a context-specific menu to appear. Different context menus will appear depending upon whether you click a [duplicate file group](#), or one of the duplicate files within a group.

Context menu functions include:

- **SmartMark** - These are extensions to our powerful SmartMark technology that are duplicate file specific. These provide a broad range of marking capabilities such as marking all files in the same group, marking files in the same (or other) folder(s), and more.

- **Open file** - Opens the currently selected file with the default Windows associated file.
- **Rename file** - Allows the selected file to be renamed.
- **Explore folder** - Opens Windows explorer and shows the folder containing the currently selected file.
- **Exclude parent folder from future searches** - Adds the parent folder to the [search exclusions listing](#).

- **Move marked item(s)** - Provides a facility for moving any currently marked files.
- **Zip marked item(s)** - Provides a facility for zipping (e.g. compressing) any currently marked files.
- **Delete marked item(s)** - Provides a facility for deleting any currently marked files.

- **Remove selected item(s) from report** - Prunes selected file entries from the report without removing them from the file system.

- **Cmd prompt here** - Opens a command prompt window within the containing folder.

- **Copy path(s) to clipboard** - Copies the full duplicate file path to the Windows clipboard.
- **Copy marked path(s) to clipboard** - Copies all marked paths to the Windows clipboard.

- **Compute file hash** - Starts the file hash calculator with the selected duplicate file paths.
- **Properties** - Displays Windows file properties window.

- **Filter search results by** - Allows search results to be filtered by various duplicate file properties.

Tip: Multiple duplicate files can be selected by holding down the Ctrl or Shift keys while clicking.

3.7 Marking Files

Before you can take action against one or more files that appears within the Duplicate File Detective result report window, you must *mark* them. You mark a file by placing a mark in the small box near the name of the files you wish to act upon.

The Marking section of the application Ribbon Bar provides a range of functions designed to assist with marking files, including [SmartMark™](#) - our special duplicate file marking assistant technology.

You can also right-click a duplicate file to gain access to numerous file and group level SmartMark functions.

After you've marked one or more duplicate files, Duplicate File Detective can further help you to perform additional processing on them. Duplicate File Detective has integrated facilities for [moving](#), [zipping](#), or [deleting](#) the duplicate files that you've marked.

See also:

[Duplicate Groups](#)
[SmartMark](#)

3.8 SmartMark

Duplicate File Detective supports a concept known as SmartMark™, allowing users to [mark](#) duplicate files using specific patterns. Marked duplicate files are then eligible for [additional processing](#) (e.g. moving, deleting, or zipping).

To use the SmartMark™ facility, click the SmartMark™ button in the application Ribbon Bar. A pop-up menu will appear, providing access to a number of duplicate file marking options.

- **Mark First File in Each Group** - Marks the first file in each duplicate group. Very useful when combined with [result report grouping](#) capabilities.
- **Mark All But First File in Each Group** - Marks all but the *first* file in each duplicate group.
- **Mark All But Last File in Each Group** - Marks all but the *last* file in each duplicate group.

- **Mark Matching File Names** - Provides a means of marking duplicate files with names matching a specific pattern.

- **Mark Oldest Files in Each Group** - Marks the oldest files in each duplicate file group.
- **Mark Newest Files in Each Group** - Marks the newest files in each duplicate file group.

- **Mark Largest Files in Each Group** - Marks the largest files in each duplicate group (using raw file size for comparison).

- **Mark Smallest Files in Each Group** - Marks the smallest files in each duplicate group.
- **Mark Longest File Names in Each Group** - Marks the longest file names in each duplicate group.
- **Mark Shortest File Names in Each Group** - Marks the shortest file names in each duplicate group.

IMPORTANT: In addition to the SmartMark capabilities outlined above, you will find *additional SmartMark features* by [right-clicking on duplicate files in the duplicate file report](#).

Tip: Some SmartMark™ functions *may not be applicable* to certain duplicate file groups. For example, if all the files within a duplicate group have the same date / time stamp, the "oldest" and "newest" smart marking option will have no effect on that group.

See also:

[Marking Files](#)

3.9 Moving, Zipping & Deleting

Duplicate File Detective allows you to act upon [marked duplicate files](#) in a number of ways. Each of these functions is accessible via the application Ribbon Bar.

- **Move** - Allows you to move marked duplicate files to another file system location.
- **Zip** - Allows you to zip (compress) marked duplicate files.
- **Delete** - Allows you to delete marked duplicate files. Please use extreme caution when exercising this option (see important note at the bottom of this page).

Selecting any of these actions will cause the **Duplicate File Manager** window to appear. This window allows you to review specifically which files will be impacted by the chosen operation. You can also still change your mind at this point and decide, for example, that you'd prefer to zip a set of duplicate files rather than delete them.

Once you've provided the appropriate move, zip, or delete criteria, click the Go button to proceed. Once the selected operations have completed, the Duplicate File Detective results report window will be refreshed to reflect any file system changes that occurred.

Note that the Duplicate Result Manager will enact or enforce whatever options are defined within the [Protection](#) tab of the global options window.

Duplicate Processing Options

The following options are available when moving, zipping, or deleting files using the Duplicate File Manager window:

- **When zipping duplicates, delete original files after adding them to zip file** - When enabled, this option will delete each file that is added to the target zip file. Note that this option will honor the "Always delete to Windows Recycle Bin" switch explained below.
- **Always delete to Windows Recycle Bin** - Uses Windows shell facilities to delete files directly into the Recycle Bin. Note that since the Windows shell itself cannot handle extremely long paths (those exceeding 255 characters in length), this option has the same limitation.
- **When moving duplicates, retain folder structure** - When enabled, the folder structure of the source files will be re-created as duplicate files are moved into the target directory. **Caution: Use of this feature within deeply nested directory structures can create extremely long paths within the target**

directory (which, in turn, may be difficult for Windows Explorer and other shell mechanisms to manipulate effectively).

- **When moving duplicates, rename files in order to avoid conflicts** - When enabled, any computed target file path that already exists will be renamed to avoid conflicts. When in effect, the resulting file names will be appended with incrementing numerical identifiers (as needed).
- **Replace moved or deleted duplicates with link to unmarked sibling** - When enabled, any action that triggers the deletion or movement of a source file will trigger the creation of a link in its place. The resulting link will resolve to the first unmarked file within the same duplicate group. If no such duplicate target exists (e.g. all entries within the same group are marked), this option will have no effect. The following types of links are supported:
 - **Shell shortcuts** - Creates .lnk files in place of moved or deleted duplicate files. Such shortcuts can be resolved by most parts of the Windows shell (including Explorer), but some third-party applications may not follow them.
 - **Hard links** - Creates hard links in place of moved or deleted duplicate files. Hard links are generally transparent to users and most applications. **The primary limitation of hard links is that they cannot span volumes.** In other words, you cannot hard link a file that resides on one physical volume to a file that resides on another.
 - **Symbolic links** - Creates symbolic links in place of moved or deleted duplicate files. **Supported only on Windows Vista / Server 2008 and later.** Use of this feature also requires that Duplicate File Detective be run from an administrative account "as an administrator" (if needed, right-click on your Duplicate File Detective start menu shortcut and choose "Run as administrator"). Such links can be resolved by most applications running on Vista / Server 2008 or later, and can span volumes (they can even resolve to paths on the network).
- **Halt processing upon occurrence of first duplicate file processing error** - By default, errors that occur during duplicate file processing will be accumulated and reported when file processing completes. When this option is engaged, all processing is halted upon the first occurrence of any such error.

Tips for successful duplicate file processing:

- Always back up source files before enacting duplicate file processing that might affect them.
- Be sure that you always understand the potential impact of moving or deleting items from any file system.
- Use the option to delete to the Windows Recycle Bin whenever possible, as it adds an additional safeguard.
- Pay close attention to any warning messages that appear prior to the execution of duplicate file processing.

Important Notice!

Before you move, zip, or delete any files on your system, please be sure that you understand the potential impact of your actions. Removing critical system files, for example, can render your system inoperable. Make sure you understand the purpose of any file system object before altering or removing it. **Never assume that a file can be safely archived or removed simply because it is a duplicate.**

3.10 Exporting

Once Duplicate File Detective has finished generating the results of a duplicate file search, the Export button in the Results Processing section of the application Ribbon Bar will be enabled.

Duplicate result reports can be exported in three different formats:

- **HTML** - HyperText Markup Language. This option also copies the "style.css" cascading style sheet file from the "template" sub-folder of your Duplicate File Detective installation directory to the file export target path (and links it to the generated HTML file). You can customize the appearance of HTML exports by modifying the "style.css" template file (knowledge of CSS syntax required).
- **CSV** - Comma Separated Values. The resulting report can be opened directly within most spreadsheet applications such as Microsoft Excel.
- **XML** - eXtensible Markup Language, consumable by a large array of third-party applications. Reports exported in XML format can be [re-imported](#) later. Note that XML export files also contain duplicate report summary data (such as report run time, files scanned, etc.) so that the [Summary Report](#) function will work correctly when data is re-imported.

You can also select precisely which duplicate file report columns you wish to export. Check the box adjacent to the name of each column you wish to include in the export file.

Other export options:

- **Open report after generation completes** - Will attempt to open the target file once the export process has completed. Note that processes uses the default Windows operating system file association.
- **Warn if target file path already exists** - When this setting is disabled, the export target file will always be overwritten, without prompting.
- **Close this window when export completes** - Close the export window when the export process completes.
- **Merge CSS template into HTML export file** - Merges the contents of the CSS export template directly into the HTML export file. When disabled, the CSS file will be placed adjacent the HTML file and linked to it.

To export a duplicate results report, simply select the export format, provide a destination file path, and click the Export button.

Note: Report export capabilities are also accessible through the [command line](#) interface.

See also:

[Importing](#)

3.11 Importing

If you [export](#) duplicate file results report in XML format, you can later re-import them via File | Import XML Report.

Provide the path to the XML report export file you created previously and click the Import button. Note that for large XML data files, this process may require some time.

There is also an option to prune the report of stale duplicate entries after import completes. When enabled, this option triggers the [Autoprune](#) feature, which in turn will compare imported duplicate search results with the file system (and remove those that are no longer present).

Important: When report results are re-imported, the project settings (e.g. search paths, exclusions, filter criteria, etc.) associated with the previous report will not be restored automatically. If you need the project and data to match, it is recommended that you load a project file first, then import the corresponding XML data file.

See also:

[Exporting](#)

3.12 Summary Report

The duplicate file Summary Report is designed to give you a quick snapshot of search results. The summary report is accessed through the Results Processing section of the application Ribbon Bar.

The report contains the following information:

- **Report date/time** - The date and time that report generation completed.
- **Search paths** - Listing of paths searched for duplicates.
- **Excluded paths** - Listing of excluded paths, if any.
- **Report run elapsed time** - The amount of time required to generate the duplicate file report.
- **Duplicate files** - Count of duplicate files found, based upon [project search filter criteria](#).
- **Files scanned** - The number of files actually scanned, based upon project filter criteria.
- **Files skipped** - Number of files skipped, based upon project filter criteria.
- **Folders scanned** - Number of folders scanned.
- **Folders skipped** - Number of folders skipped.

If an entire folder is skipped during scanned (perhaps because it was directly [excluded](#), or indirectly through [filter options](#)), the files it contains will **not** be counted as "skipped" files. Skipped files are only those directly encountered during the initial file system scan, but not used for duplicate file comparison purposes for some reason (e.g. filter criteria settings, etc.).

Note that the contents of the summary report window will be restored when XML data files are [imported](#).

The summary window also provides an option to automatically show itself whenever a duplicate scan (or import) completes. You'll find a checkbox for controlling this option on the Summary Report window itself.

3.13 Autoprune

The Autoprune function compares current duplicate file search report entries against the file system to determine if they're still present. If not, the associated report entry is removed (pruned).

By default, Autoprune is triggered automatically after the [Import](#) function is used.

Autoprune is also launched automatically after duplicate files are processed with Duplicate File Detective (using the built-in [moving, zipping & deleting](#) features).

3.14 Printing

Once Duplicate File Detective has finished generating the results of a duplicate file search, the Print button in the application Ribbon Bar will be enabled.

Clicking the Print button will cause a window to appear that allows you to choose the destination printer, to which you can then print the current duplicate report results.

Duplicate File Detective also provides useful print preview and setup functions.

See also:

[Result Report](#)

3.15 Result Filtering

Duplicate search report results can be filtered, so that you see only a specific set of duplicate files that match the specified filter text.

Once the duplicate search process completes (or results are imported), the Results Filtering section of the application ribbon bar (contained with the View tab) will become enabled.

To filter the on-screen [duplicate results](#), type some text into the filter input box and press enter (or select a previously used entry, if applicable, using the adjacent drop arrow). The duplicate results report will be updated to show only entries that contain the filter text you entered.

By default, the filter text entered in the step above will be compared to the filenames of the duplicates in the search results report. But you can elect to filter by other duplicate file properties as well (such as folder, extension, owner, etc.). Use the drop-down box adjacent the filter input box to choose how filtering will be applied.

Filtering also offers two separate modes of operation, one that shows duplicate file group siblings and another that does not. Use the Filter Mode button to choose which mode you prefer.

Click the Clear button in the Results Filtering section of the application Ribbon Bar (in the View tab) to remove the currently applied filter.

3.16 Image Preview

The Image Preview docking panel provides a visual reference point when using Duplicate File Detective to identify duplicate image files.

To preview an image file, select it within the main report detail view and the image preview panel will display a thumbnail version of it. Like any docking panel within Duplicate File Detective, the image preview window can be resized, re-positioned, or docked / undocked at your discretion (just drag the panel's window title where you want it and release it).

Supported image preview file formats are: BMP, EMF, GIF, ICO, JPEG, PNG, TIFF, and WMF.

Note: The image preview window will display thumbnails for supported image files with a size less than or equal to that defined in the [advanced tab](#) of the [preferences](#) window.

3.17 Folder View

The Folder View docking panel provides a hierarchical view of folder structures involved in the current duplicate search [result report](#), and also provides a means of controlling their visibility.

When enabled, the Folder View is automatically populated at the completion of a duplicate search process or an [XML import](#). The folder structure shown in the Folder View is always deduced from the

records available in the duplicate search results view, and will be updated automatically after any pruning operations occur (through manual removal of records or use of [AutoPrune](#)).

Not every folder shown in the Folder View tree will always directly contain duplicate files. Some nodes are present to better represent the overall folder hierarchy, as well as allow folder tree check boxes to operate in hierarchical manner (see description of Smart Checkboxes feature below). By default, folders that **do directly contain duplicates are shown in bold** and can be quickly navigated through use of the "Next/Previous Folder With Duplicates" commands, described below.

Folder Filtering

The check boxes adjacent each folder name within the Folder View tree indicates whether the corresponding duplicate search results report entries are visible or not. In other words, if you clear the check box associated with a folder in the Folder View tree, it will cause all search result entries in that folder to become hidden.

The effect of folder level filtering is always shown near the bottom of the Folder View docking panel. When no folder tree nodes are unchecked, it will display the text "Hiding: None". When one or more folder tree nodes are unchecked, this area will display the total number of files hidden along with the number of folders that contain them.

Note that if the Folder View docking panel is closed, any folder level filtering currently in effect will be reset.

Folder View Commands

The Folder View panel has its own toolbar, which exposes the following commands:

- **Check All** - Checks all folder tree nodes, effectively resetting any filtering current being applied.
- **Uncheck All** - Unchecks all folder tree nodes, effectively hiding all duplicate file entries in all folders.
- **Next Folder With Duplicates** - Finds and selects the next folder in the tree that contains duplicate files. Uses the current selected folder tree node as the starting point, or the root node if no selection currently exists.
- **Previous Folder With Duplicates** - Find and selects the previous folder in the tree that contains duplicate files.
- **Folder View Tree** - Exposes a pop-up menu with its own set of commands:
 - **Expand All Folders** - Expands the entire folder view tree to full depth.
 - **Collapse All Folders** - Collapse all folder view nodes.
 - **Invert All Checks** - Inverts the state of all tree node check boxes. This function is only enabled if the Smart Checkboxes feature is turned **off**.
 - **Bold Folder Names Containing Duplicates** - Shows the names of folders that directly contain duplicate files in bold.
 - **Enable Smart Checkboxes** - When engaged, causes check boxes to operate in a hierarchical manner. For example, checking a parent folder node will cause all of its children to be checked as well. Note this feature is incompatible with some other commands such as "Invert All Checks".
 - **Rebuild Folder Tree** - Causes the folder tree structure to be re-built based upon current duplicate search results.
 - **Disable Folder View** - When engaged, causes Folder View to become completely disabled, bypassing automatic folder tree population. You may wish to disable the Folder View (for performance reasons) when managing very large numbers of duplicates.

A subset of these commands is also available on a per-folder basis. To access these commands, right-

click a node within the folder tree:

- **Find First Duplicate in Folder** - Finds and selects the first duplicate file (in the duplicate file results view) that is contained within the selected folder.
- **Mark Duplicate Files in Folder** - Marks all duplicate files contained within the selected folder.
- **Unmark Duplicate Files in Folder** - Unmarks all duplicate files contained within the selected folder.
- **Mark Duplicate Files in Folder and All Subfolders** - Marks all duplicate files contained within the selected folder and any of its child folders.
- **Unmark Duplicate Files in Folder and All Subfolders** - Unmarks all duplicate files contained within the selected folder and any of its child folders.
- **Check (Show) Folder and All Subfolders** - Toggles visibility checkbox for current folder and its children, effectively showing them.
- **Uncheck (Hide) Folder and All Subfolders** - Clears visibility checkbox for current folder and its children, effectively hiding them.
- **Expand All Subfolders** - Expands all children of the currently selected folder.
- **Collapse All Subfolders** - Collapses all children of the currently selected folder.
- **Check All But Selected Folder** - Toggles visibility checkbox for all but the currently selected folder, effectively showing them. *This function is available only if Smart Checkboxes (see description above) are disabled.*
- **Uncheck All But Selected Folder** - Clears visibility checkbox for all but the currently selected folder, effectively hiding them. *This function is available only if Smart Checkboxes (see description above) are disabled.*

3.18 File Properties

The File Properties docking panel appears, by default, at the lower left of the main Duplicate File Detective user interface.

To show properties for a file, simply select it in the duplicate report results listing. The File Properties panel will be updated to reflect the properties of the currently selected file.

The following information is displayed:

- **General** - Name, full path, and other basic file information.
- **Size** - File size information.
- **Date & Time** - Shows all date / time file attributes.
- **Attributes** - Shows attributes that apply to this file (e.g. read-only, hidden, etc.).
- **Version** - Shows file version information (appears only for executable file types).

The File Properties docking panel also offers several functions via its integrated toolbar:

- **Save** - Writes a copy of the file properties to a text file in the location of your choosing.
- **Copy** - Places a textual representation of the file properties on the Windows clipboard.
- **Categorize** - Breaks down file property data into a series of logical categories.
- **Alphabetize** - Shows file property data in alphabetical order.

Note: Information displayed within the File Properties docking window is read-only (e.g. modifications to file properties cannot be made from here).

4 File Types

4.1 About File Types

As a duplicate file searched progresses, Duplicate File Detective monitors and classifies all file type information that it encounters. The resulting information is shown in the File Types docking panels (located at the bottom of the main application window by default).

The File Types report views (one list view and two graphs) display the distribution of duplicate files across various types of files. If, for example, most of your duplicate files on a particular searched file system path are .GIF images, you can easily see as much.

File Type Views

- [Report view](#) - shows file type information in tabular form
- [Graph by count](#) - shows distribution of file types by count
- [Graph by size](#) - shows distribution of file types by file size

4.2 File Types Report

The File Types report view shows the distribution of scanned files (including duplicates) across various file types. Available columns include:

- **File type** - The type of file in question (e.g. GIF Image)
- **Extension** - The associated file extension (e.g. .gif)
- **Files** - The number of files of this type encountered during searching
- **Size Files** - The total size of all files of specified type
- **Duplicates** - The number of duplicates of this type encountered during searching
- **Size Dups** - The total size of duplicate files of specified type
- **% Dups** - Percentage files found of specified type

Toolbar Functions

- **Print** - Print the file types report
- **Save** - Output the file types report data as comma separated values (e.g. CSV format)
- **Show all** - Shows all file types encountered during duplicate scanning, even if no duplicates were found (by default, only file types for which duplicates were found are shown)

Context Menu Functions

You can right-click on any file type row and either mark report entries and:

- **Mark report entries of this type** - Causes all associated file entries in the main application report view to be marked (the check box for the file will be checked)
- **Filter report view by this type** - Implements a main application report view filter in the form of *.ext (only file of the associated type are shown in the main report view)

4.3 File Types Graphs

The File Types Graphs show the "Top N" distribution of scanned and duplicate files by count and file size.

You can control how many entries are displayed in the File Types Graphs within the [Global Preferences](#) window (see the [File Types](#) tab).

Toolbar Functions

- **Print** - Displays graph print preview window, which allows you to resize the graph output, etc.
- **Save** - Allows you to save the graph image in Bitmap, JPEG, or EPS format
- **Copy** - Copies the graph image to the Windows clipboard (you can then paste the image into another application)

4.4 File Type Researcher

The File Extension Researcher tool helps you to determine the potential function (or application association) of a given file extension (such as ".PDF" or ".DOC").

It's important to understand that there is no central authority governing the use of file extensions by software applications. So even though many file extensions are predominantly associated with a single application (e.g. ".doc" = Microsoft Word Document), such direct associations are absolutely not guaranteed. Further, you will often encounter file extensions for which the operating system has no known software association. The File Extension Researcher helps you to determine the potential origin of such files.

To use the File Extension Researcher, enter a file extension (such as ".pdf" or ".doc") and click the Go button. You will be presented with two distinct pieces of information:

- **Shell Association** - The operating system shell has special, intrinsic knowledge of some file types and their application associations. The shell uses this knowledge to (among other things) open files with specific extensions with a default application.
- **All Potential Associations** - This is a more complete listing of functional / application associations. This list will often yield results even when the Windows shell has no knowledge of a particular file extension.

The "All Potential Associations" listing shown within the File Extension Researcher pulls information from a file extension database contained within Duplicate File Detective. This database contains thousands of mappings between file extensions and application / functional associations, but it can never be entirely complete (because there's no central registry for file extensions from which to cull such data).

You can also click the Research Online button, which will launch your web browser and search for details (and other potential associations) on the file extension you entered.

Finally, you can browse the entire file extension database contained within Duplicate File Detective by clicking the "View All Extensions" button. The resulting window will allow you to browse known file extensions by their first letter, or by numeric / symbolic grouping.

5 Preferences

5.1 About Preferences

In Duplicate File Detective, application Preferences are those not directly associated with the current [project](#). Preferences are not saved and loaded as part of a project, and are in effect regardless of what project is currently loaded.

5.2 General

General Options

- **Re-open last project when program starts** - Indicates whether the last [project](#) will be loaded automatically upon startup. Enabled by default.
- **Follow file system re-parse points** - Causes file system scanning to follow file system reparse points (such as junctions). Disabled by default.
- **Enable typing auto-complete within application drop-down boxes** - Enabling this switch allows the system to automatically complete text being entered into various drop-down combo boxes used throughout the system. Enabled by default.
- **When filtering duplicate search results, show siblings of matching items** - When enabled, filtering items within the duplicate search result view will also cause all members of the same duplicate file group to be shown.

Prompts & Notifications

- **Never ask if changes to the current project should be saved** - Whenever a project is closed (e.g. when shutting down Duplicate File Detective or starting a new project), you'll be prompted to save any unsaved changes. Use this setting to avoid such prompts. Disabled by default.
- **Show warning when result report filtering is applied through indirect means** - Some right-click menu functions may apply filtering to the current duplicate search result view. When this option is enabled, users will be notified when such filtering has been applied.
- **Show desktop notification window when duplicate scan completes** - Causes a small notification window to appear in the corner of the Windows desktop when a duplicate scan completes. Notifications are never shown if the current duplicate file scan is canceled by the user.
 - **Only if one or more errors occur during scanning** - Causes the notification window mentioned above to appear only if errors occurred during the duplicate file scanning process.

5.3 Protection

The Protection tab within the Global Preferences window provides an additional layer of safety to duplicate file management operations executed through Duplicate File Detective.

Warnings

Duplicate File Detective provides additional protection by *warning* you if you're about to move or delete a file that violates certain safety constraints. For example, if you produce a duplicate file report, [mark](#) one

or more duplicate files, and then click the [Delete](#) button in the Results Processing section of the application Ribbon Bar, the resulting [Duplicate Result Manager](#) window will use the Protection options to determine whether or not you're doing something potentially unsafe.

The protection warning options available are generally self-explanatory. Such warnings, if applicable, will be issued after you click the Go button in the [Duplicate Result Manager](#) window.

Protected File Types

Protected file types are represented by an editable list of file extensions, each of which represents a file type that probably should not be manipulated by any duplicate file management process.

Note: By default, the [filename matching](#) section of the [Search Filtering](#) panel will exclude protected file types from duplicate file search operations. Disable the appropriate switch if you wish to search for duplicates of these file types.

5.4 Results Report

File Attribute Colors

Individual file entries in the duplicate [results report window](#) are color-coded according to their file attributes. For example, by default compressed files are displayed in blue and system files are displayed in maroon.

Here, you can change the colors used to display normal, compressed, hidden, read-only, or system files.

Note: When a file has multiple attributes that would affect the display color, the first matching color (in the order specified above) is used. To see all the attributes that apply to a specific file, simply select the file in the report list and refer to the [File Properties](#) docking panel.

Search Results Report

- **Show filename of marked items with strike-through font** - When enabled, any marked report item filenames will be displayed with a line drawn through them.
- **Lighten the color of marked item rows** - When enabled, the color of any marked report items will appear lighter than those that are unmarked. You can also choose the percentage of lightening to apply.
- **Show time (along with date) in date-based report columns** - Shows both date and time in date report fields such as "last modified", "created", and "last accessed". By default, only the date is shown (without the time).
- **Show tooltips when hovering over report fields** - When enabled, tooltips will appear when the user's mouse hovers over any part of a duplicate search report field. Note that these tooltips may contain additional information about the duplicate file entry in question (such as duplicate group summary information, etc.).
- **Automatically mark all but the first of each duplicate file group** - When enabled, applies a default set of file markings when a duplicate file search process completes and results are shown.
- **Hash encoding type** - Controls how file hash values are displayed within the duplicate search result report.

5.5 Update Checking

Proxy Options

Options presented in the Updates tab of the Global Preferences window allow you to perform update checks via an HTTP proxy server.

To use an HTTP proxy server when performing update checks, select the "Use HTTP proxy server" check box and provide a proxy port and address (either a host name or an IP address).

Other Options

- **Automatically check for updates** - When enabled, Duplicate File Detective will check for updates whenever it is started. This will occur in the background, and a small tray notification area window will be shown if a newer version of the application is available.

5.6 New Projects

When a new [project](#) is created within Duplicate File Detective, a default set of search and exclusion paths are added to the project.

Default Search Paths

- **Add fixed disk drives** - Enabled by default.
- **Add remote (mapped) drives** - Disabled by default.
- **Add removable drives** - Disabled by default.

Default Exclusions

- **Exclude Windows folder** - Where OS and other critical files are kept (e.g. C:\windows).
- **Exclude Program Files folder** - Where installed programs and their supportive files reside.
- **Exclude Common Program Files folder** - As above, only for all users.
- **Exclude Windows System folder** - The Windows system folder (e.g. C:\Windows\System).
- **Exclude Temporary Internet files folder** - The file system location in which Internet Explorer places temporary files.

All of the above exclusions are enabled by default.

5.7 File Types

The File Types tab of the Global Preferences window allows you to alter the behavior of the [File Types](#) report windows.

Graphs

- **Show stacked graphs** - Shows graph values (scanned / duplicate files) stacked upon one another, rather than side-by-side. Disabled by default.
- **Show marks** - Shows graph markers indicating the precise value of graph items. Disabled by default.
- **Show 3D graphs** - Disable this setting to "flatten" the graphs. Enabled by default.
- **Show top N file types** - Indicates how many "Top N" graph items should be displayed. Default value is 10.

5.8 Logging

Duplicate File Detective can log various details about duplicate file search operations to the file system.

This logging facility was designed to assist users in understanding and/or troubleshooting duplicate search processes and results. It should be used only as needed, and otherwise left disabled (please see additional comments below).

Events to Log

- **File found** - Logs the path of files found by initial search pass.
- **File skipped** - Logs the path of files skipped during initial search pass. Includes information about why each file was skipped.
- **Folder scanned** - Logs the path of each folder searched.
- **Folder skipped** - Logs the path of each folder skipped. Includes information about why each folder was skipped.
- **File compared** - Logs full path of comparison candidates. Includes a description of the file comparison result.

Important: Use of logging can degrade performance and consume large amounts of disk space. Logging should be enabled only for short durations in support of troubleshooting operations, etc. All log files are written in CSV format to the logging directory specified above, and are overwritten between project runs (even when logging is disabled).

Log File Directory

Allows users to configure where log files are created and stored.

5.9 Advanced

This tab of the Global Preferences window provides access to advanced Duplicate File Detective settings.

Performance

- **Suppress retrieval of file ownership information** - File ownership data is automatically retrieved from duplicate files for display within the duplicate report (upon search completion). Extraction of this information can slow duplicate search performance, especially when scanning remote (network) paths. If you don't care about file ownership, you may engage this switch to suppress its collection (file ownership fields will always be displayed as "unknown").
- **Use low-priority disk I/O for scanning & hashing** - When enabled, allows operating system to prioritize other disk I/O requests over that of Duplicate File Detective. Requires re-start to take affect.
- **Number of threads used during file content hashing operations**
 - **Single threaded** - Recommended for local storage devices, which will not generally benefit from multi-threaded disk I/O. Default.
 - **One thread per CPU core** - Recommended for network attached storage devices. May also be useful when scanning multiple discrete local volumes.
 - **Use this specific number of threads** - Allows for fine-grained control over thread allocation.

Tip: Multiple file content hashing threads are generally *most* effective at speeding up duplicate file comparison operations when the current project is configured to scan multiple, discrete volumes (e.g.

paths spanning multiple physical disks or multiple network paths). Using large numbers of threads to hash files located on a single physical volume will usually only serve to *reduce* performance, and is not recommended.

Miscellaneous

- **Maximum image preview file size (in KB)** - Image files selected within the report detail view will be previewed only if their file size is less than what's defined here. Default value is 102,400 KB (100 MB).
- **Enable Unicode support when creating zip archives** - Allows for the creation of zip files that support Unicode file paths. Before engaging this option, be sure to confirm that any archive processing applications you might use to open these files also support Unicode.

6 Advanced Features

6.1 Customizing

Duplicate File Detective has a huge variety of functional and user interface customization options.

Ribbon Bar

You can customize the application Ribbon Bar by clicking the round button with the Duplicate File Detective icon, then selecting Customize. The resulting window provides the ability to customize application commands, keyboard shortcuts, and more.

Docking Panels

Any of the panels that appear by default on the right and left of the main application window can be docked, undocked, hidden, or closed entirely. Simply click the caption bar of one of the panels (the *Filter Criteria* panel, for example) and drag the mouse to move the panel. A series of guides will appear that help you to visualize and access the available docking positions. You can even dock a panel *inside* of another docking panel, in which case it appears as an additional tab within the panel frame.

The visibility of Docking Panels can also be manipulated via the View tab of the Ribbon Bar.

Report Options

Right-clicking the duplicate file report column header will allow you to add and remove report columns, alter column alignment, and more. Report columns can also be managed via the View | Report Field Chooser menu item - when the field chooser window appears, you can drag report columns from it and drop them onto the report detail area (the reverse can also be used to remove them).

Duplicate file report columns can also be re-ordered simply by dragging and dropping.

Additional report options can be found by clicking the main View menu, which contains options for controlling how the report displays file sizes, the report grid style, how grouping is performed, shading of report group headers, and more.

Global Preferences

Some report customization options (such as [attribute colors](#)) are located within the global preferences window, accessible via the main Tools menu.

Application Theme

The drop-down selector near the upper right of the main window frame provides a choice of application themes. Each theme affects the application Ribbon Bar, as well as docking panels, etc.

6.2 Command Line Options

Duplicate File Detective provides a command-line interface that you can use to schedule project execution or integrate with an existing file management process.

Command line instructions (known as *arguments*) are passed in using a specific syntax which must be followed carefully. Here's an example:

```
dfd.exe /project:"c:\temp\test-project.dfd" /html:"c:\temp\export.html" /exit
```

Note that the arguments passed to the command line actions are surrounded by quotes; this is a requirement, as it allows the argument to contain spaces and still be parsed correctly.

You can also access the Duplicate File Detective executable using an absolute path, which is generally what you'll want to do when creating batch files, etc. Example:

```
"C:\Program Files\Duplicate File Detective\dfd.exe" /project:"c:\temp\test-project.dfd"
```

Notice how, in the example above, the path to the Duplicate File Detective executable is also contained in quotes. This helps to ensure that spaces in the path name are handled correctly by the Windows command interpreter.

Important: Use of command line options within Duplicate File Detective is [project-oriented](#). You execute a project by providing its name to the Duplicate File Detective program executable. You must create and save a project to disk before you can execute it in this fashion. This is a very powerful concept which effectively provides you with access to *every project setting* from the command line.

Available Command Line Switches

- **/project** - provides the full path to the Duplicate File Detective project file to execute. This flag is **required**; without it, any other command line options are ignored.
- **/html** - [Exports](#) the duplicate file results in HTML format when project execution completes.
- **/csv** - [Exports](#) the duplicate file results in CSV (comma separated values) format when project execution completes.
- **/xml** - [Exports](#) the duplicate file results in XML format when project execution completes.
- **/exit** - Causes Duplicate File Detective to terminate once all other command line options have been processed.
- **/date** - Appends the current date to the specified export filename; this applies only when using the `"/html"`, `"/csv"`, or `"/xml"` flags. Allows for unique daily export filenames within the same parent folder.

Tip: To schedule the execution of a Duplicate File Detective project, first create a batch (.bat) file that calls the Duplicate File Detective executable with the correct command line arguments. Then use the integrated Windows scheduler to invoke the batch file on a scheduled basis.

6.3 File Hash Calculator

Duplicate File Detective provides a simple File Hash Calculator that you can use to compute the checksum of one or more files.

Using the File Hash Calculator is a good way to get familiar with [file checksum](#) concepts, and can also provide you with a rough sense of the time differences required to perform each type of hashing computation. This information can help you make better decisions on what types of [file matching](#) options to use in your projects.

To use the File Hash Calculator, select the [hash types](#) you wish to have produced, enter one or more file paths, and click the Compute File Hashes button. The file hash results area will show the requested hash type for each file.

6.4 Error Log

If Duplicate File Detective encounters any non-fatal errors during the duplicate file searching process, it will record them in its internal Error Log.

When the execution of a project completes, if the Event Log has any entries, a special notification will appear in the lower right hand corner of your Windows desktop. The notification will alert you to the presence of the warnings or errors accumulated in the Error Log, and allow you to access them.

You can also access the Error Log via the Tools section of the application Ribbon Bar. It will only be enabled if some type of error occurred during duplicate file search operations.

6.5 Managing File Type Groups

Clicking the Presets button within the [Filename masks](#) section of the [Filter Criteria](#) docking panel will reveal the File Type Groups selection window. Double-clicking any entry in this list will cause the Filename masks drop-down box to be populated with the associated file extensions.

You can also manage these preset values by clicking the Manage button within the File Type Groups window. Doing so will reveal a window through which you can add, replace, and delete file type groups.

To add a new file type group, fill in the group name and file mask fields, then click the Add button. To replace an existing entry, select it in the top-level list, modify the entry according to your needs, and click the Replace button. Finally, you can delete an entry by selecting it in the list and clicking the Delete button.

Clicking the Save button at the bottom of the screen will cause your changes to be committed. If you cancel rather than save, your edits will be discarded.

Index

- A -

Advanced Options 12
Advanced Settings 30
Attribute Color Settings 28

- B -

Buy 4

- C -

Calculator 33
Checksums 15
Command Line Options 32
Content Hashing 12
Context Menu 16
Customizing 31

- D -

Date Filtering 11
Default Project Settings 29
Deleting Duplicate Files 18
Duplicate Groups 15
Duplicate Results 14

- E -

Error Log 33
Event Log 33
Exclusions 8
Exporting Duplicate Data 19

- F -

File Checksums 15
File Hash Calculator 33
File Matching 12
File Properties 24
File Type Groups 33
File Type Settings 29

File Types 25
File Types Graphs 26
File Types Report 25
Filename Masks 9
Filter by File Date 11
Filter by File Name 9
Filter by File Size 12
Filtering 9
Filtering Search Results 22

- G -

General Preferences 27
Global Preferences 27
Groups 15

- H -

Hash Calculator 33
Hashes 12

- I -

Image Preview 22
Importing Duplicate Data 20
Introduction 4

- M -

Managing Projects 6
Marking Duplicates 17
Moving Duplicate Files 18

- N -

New Project Settings 29

- P -

Preferences 27
Preview 22
Print Preview 21
Printing 21
Progress 14
Project Defaults 29
Project Wizard 6

Projects Defined 5
Properties 24
Protection Settings 27
Purchase 4

- R -

Registration 5
Result Filtering 22
Result Report 14

- S -

Scheduling 32
Search Paths 7
Size Filtering 12
Smart Marking 17
SmartMark 17
Sorting 15
Starting 4
Summary Report 21

- T -

Thumbnails 22

- U -

Update Checker Settings 29

- Z -

Ziping Duplicate Files 18