

# FILE RENAMING, DE-DUPLICATING, & FILE TRANSFER

---

Andrew Glass and Lauren Gaylord  
INF 389E Intro to Records Management  
October 20, 2014

# Outline & Goals

- Introduce key terms
- Relate concepts to use in records management
- Application to DIRKS projects
- Summary of tools
  - File Renaming
  - De-Duplicating
  - File Transfer
- Conclusions & Implications
- Resources

# Key Terms

- **Batch:** “Batch processing is the execution of a series of programs ("jobs") on a computer without manual intervention.” [http://en.wikipedia.org/wiki/Batch\\_processing](http://en.wikipedia.org/wiki/Batch_processing)
- **De-Duplicating:** “a specialized data compression technique for eliminating duplicate copies of repeating data.”  
[http://en.wikipedia.org/wiki/Data\\_deduplication](http://en.wikipedia.org/wiki/Data_deduplication)
- **Batch Renaming:** “Batch renaming is a form of batch processing used to rename multiple computer files and folders in an automated fashion”  
[http://en.wikipedia.org/wiki/Batch\\_renaming](http://en.wikipedia.org/wiki/Batch_renaming)
- **File Transfer:** “File transfer is the movement of one or more files from one location to another.”  
<http://searchnetworking.techtarget.com/definition/file-transfer>

# Application to Records Management

- Aid in fulfilling the characteristics of a Records System outlined by ISO 15489-1:2001: Reliability, Integrity, Compliance, Comprehensiveness, Systematic
- Can improve “efficiency and effectiveness” of electronic information systems by creating best practices for managing large numbers of files
- Can help manage two major problems of bad recordkeeping systems: poor file naming conventions and duplicate files
- Can transfer files securely within organization or to other organizations (e.g. archives), without loss of metadata

# Application to DIRKS

## **Step A: Preliminary Investigation**

De-duplication tools can be useful in preliminary analysis of subject's recordkeeping system in determining the number of duplicate files it contains.

## **Step D: Assessment of existing systems**

As with Step A, these tools can help assess the existing system where bad practices, including file names and duplicate files, are present.

# Application to DIRKS

## **Step E: Identification of strategies for recordkeeping**

These might include instituting standard naming practices. Removing duplicates. Moving files to a different location.

## **Step F: Design of a recordkeeping system**

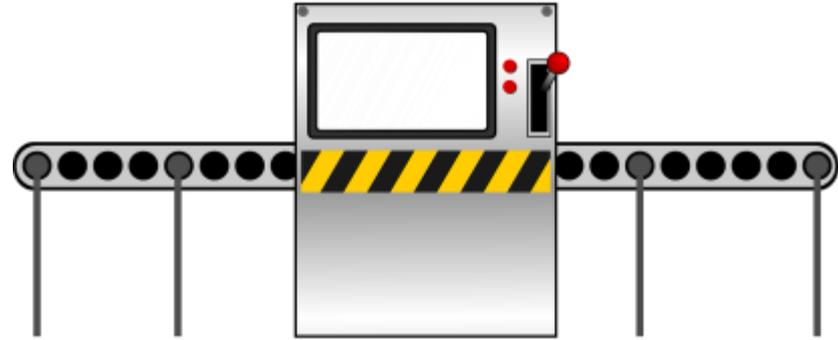
With those strategies from Step E in mind, use of the applications in this survey will be extremely useful.

## **Step G: Implementation of a recordkeeping system**

Even though we are not completing this step in class, subjects may want to implement recommendations using these tools.

# File Renaming

1. Automator (Mac)
2. Image Renaming (Windows)
3. A Better Finder Rename (Mac)
4. Siren (Windows)
5. Name Mangler (Mac)
6. RenPhoric (Windows)

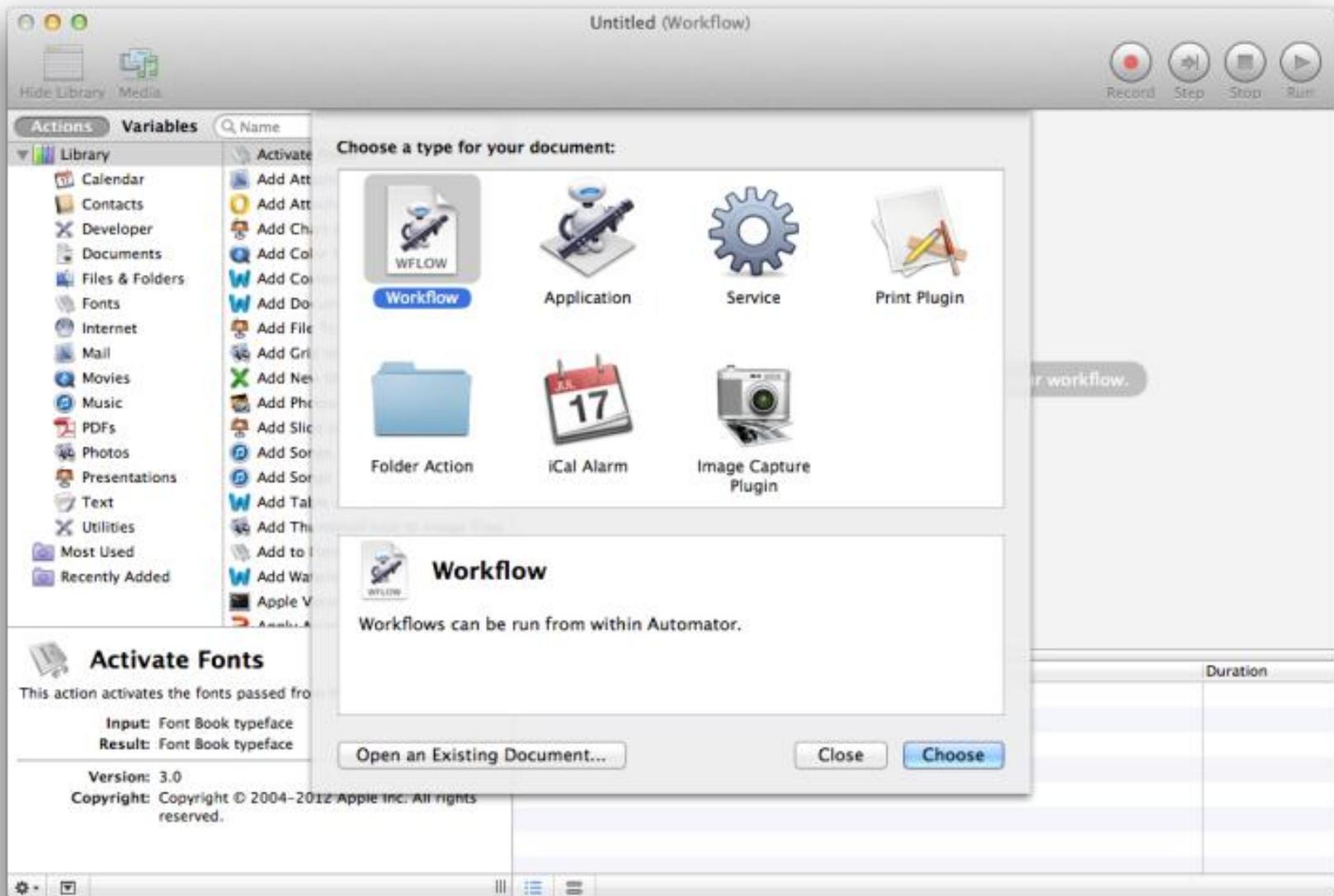


# Automator

- Operating System: Mac OS X
- Developer: Apple, included on system
- Price: Free
- Features:
  - “personal automation assistant” to make repetitive tasks easy
  - Easily rename files by date/time, sequence, or replacing text
  - Preview file names
- Limitations
  - Step by step process
- Download: built-in



# Opening Automator





# Running the Workflow

Untitled (Workflow)

Record Step Stop Run

Actions Variables get specified

Library

- Calendar
- Contacts
- Developer
- Documents
- Files & Folders
- Fonts
- Internet
- Mail
- Movies
- Music
- PDFs
- Photos
- Presentations
- Text
- Utilities
- Most Used
- Recently Added

Get Specified Address Book Items

Get Specified Fetch Items

Get Specified Finder Items

Get Specified iCal Items

Get Specified iPhoto Items

Get Specified iTunes Items

Get Specified Mail Items

Get Specified Movies

Get Specified Servers

Get Specified Text

Get Specified URLs

Name	Path
IMG_2809.JPG	classic-rocker > Pictures > Cats > IMG_2809.JPG
IMG_2817.JPG	classic-rocker > Pictures > Cats > IMG_2817.JPG
IMG_2818.JPG	classic-rocker > Pictures > Cats > IMG_2818.JPG
IMG_2819.JPG	classic-rocker > Pictures > Cats > IMG_2819.JPG
IMG_2820.JPG	classic-rocker > Pictures > Cats > IMG_2820.JPG
IMG_2821.JPG	classic-rocker > Pictures > Cats > IMG_2821.JPG

Add... Remove

Results Options Description

**Make Sequential**

Make Sequential

Add number to  existing item name  new name

Place number  Start numbers at

separated by   Make all numbers  digits long

Example: CatCafe\_01.xxx

Results Options Description

Log	Duration
Get Specified Finder Items completed	0.091 seconds
Make Sequential completed	0.151 seconds
Workflow completed	0.242 seconds

Workflow completed

Cats

- CatCafe\_01.JPG
- CatCafe\_02.JPG
- CatCafe\_03.JPG
- CatCafe\_04.JPG
- CatCafe\_05.JPG
- CatCafe\_06.JPG
- CatCafe\_07.JPG
- CatCafe\_08.JPG
- happycat.jpg

classic-rocker > Pictures > Pict

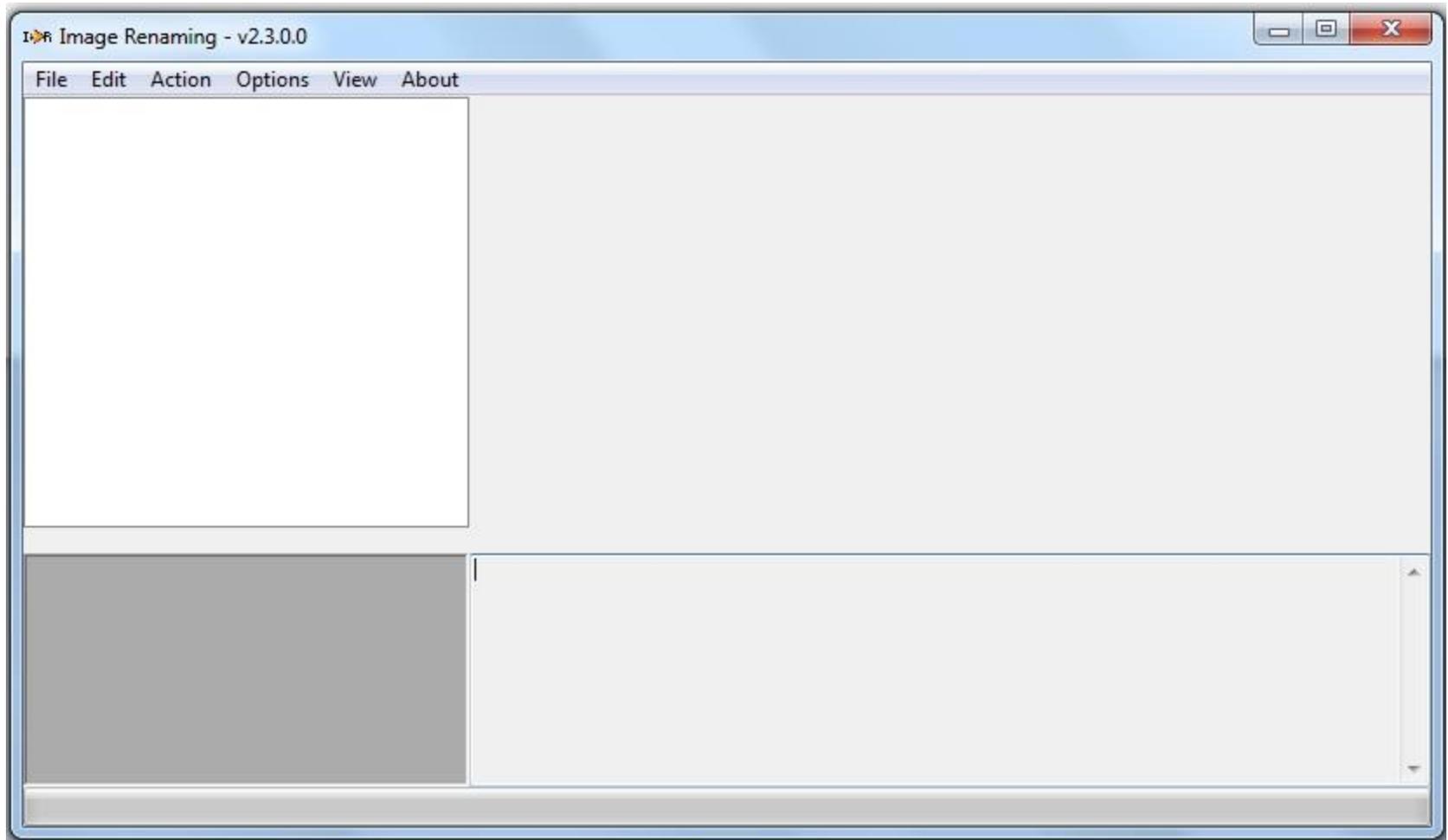
ected, 156.72 GB availa

# Image Renaming

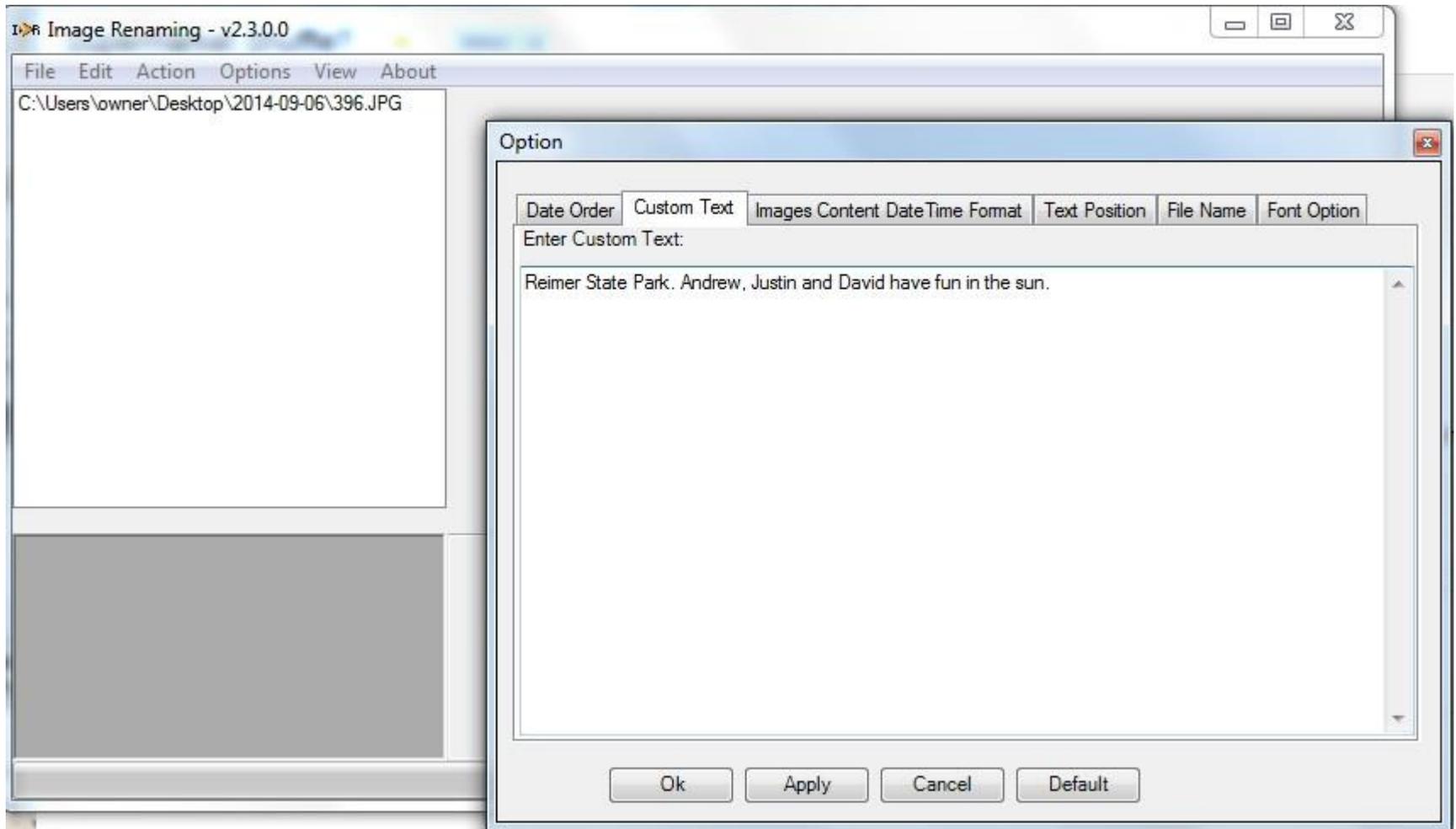


- Operating System: Windows
- Developer: Bassita
- Price: Freeware
- Features:
  - Add date to image
  - Add custom text to image
  - Add date to file name
  - Append index to file name
  - Random rename file
- Limitations:
  - Only works on photos.
  - Random rename?
- Download: <http://bassita.com/software/>

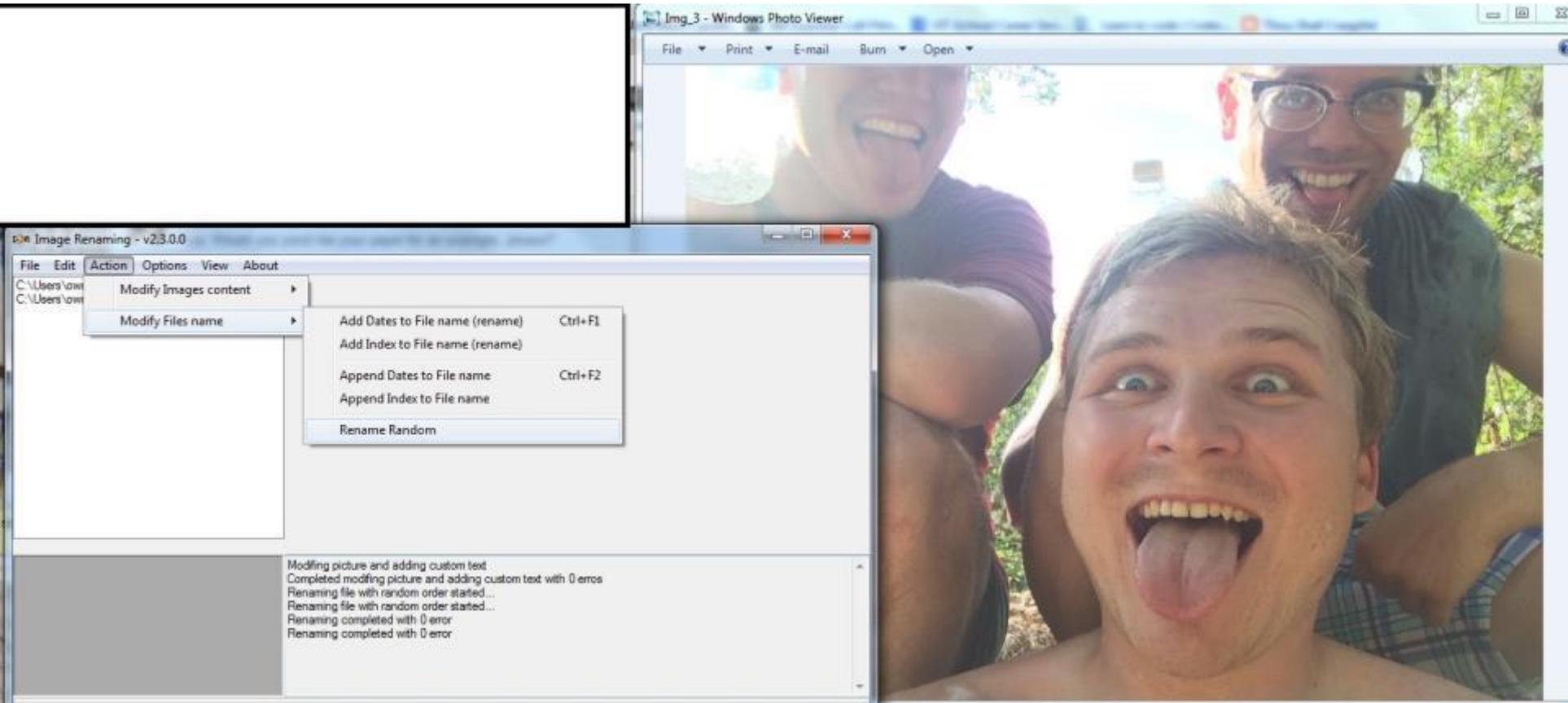
# Image Renaming Layout



# Adding Custom Text to Photo



# Random Rename



# A Better Finder Rename

- Operating System: Mac, Windows
- Developer: publicspace.net
- Price: Free trial, \$19.95 for single-user
- Features:
  - Tons of options for renaming, including tag-based
  - Multi-step renaming
  - Preview all file name changes
- Limitations:
  - Trial limits you to 10 files at a time
  - Easy to remove file extension during process
- Download:



<http://www.publicspace.net/ABetterFinderRename/index.html>

# Renaming options

Text  
Characters  
Character Positions & Ranges  
Conversions  
Truncation  
Sequence Numbers

## Roman Numerals

✓ Date & Time  
Alphabetical Sequences  
Image Dimensions  
MP3/AAC Music  
Parent Folder Name  
Path Components  
Tags  
Advanced & Special

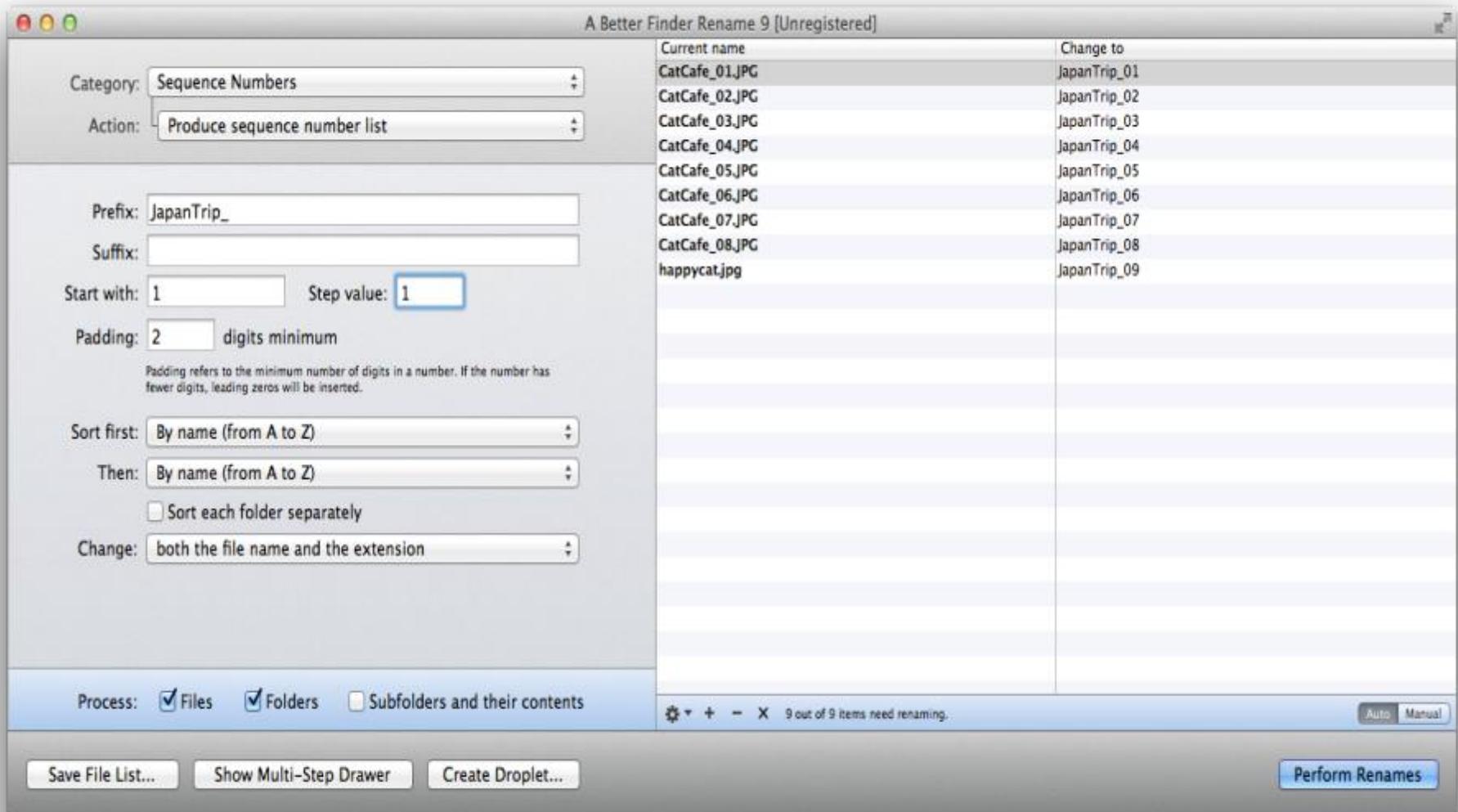
## ✓ Date Time

DateTime  
Date-Time  
Date\_Time  
Date/Time  
Date.Time  
Time Date  
TimeDate  
Time-Date  
Time\_Date  
Time/Date  
Time.Date

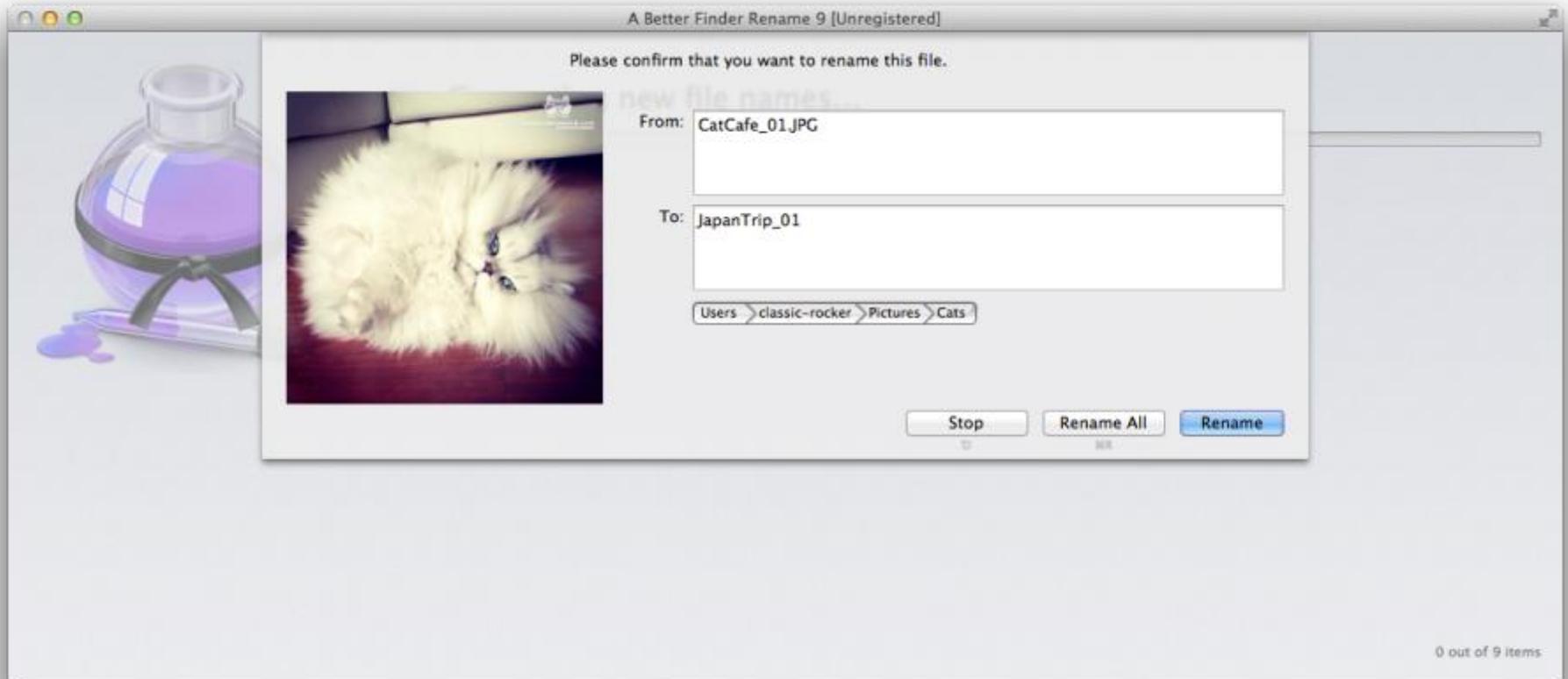
(Do not include date)

✓ MM/DD  
MM/DD/YY  
MM/DD/YYYY  
MM-DD  
MM-DD-YY  
MM-DD-YYYY  
MM.DD  
MM.DD.YY  
MM.DD.YYYY  
MM\_DD  
MM\_DD\_YY  
MM\_DD\_YYYY  
MMDD  
MMDDYY  
MMDDYYYY  
DD/MM  
DD/MM/YY  
DD/MM/YYYY  
DD-MM  
DD-MM-YY  
DD-MM-YYYY  
DD.MM  
DD.MM.YY  
DD.MM.YYYY  
DD\_MM  
DD\_MM\_YY  
DD\_MM\_YYYY  
DDMM  
DDMMYY  
DDMMYYYY  
YY/MM/DD  
YYYY/MM/DD  
vv-mm-dd

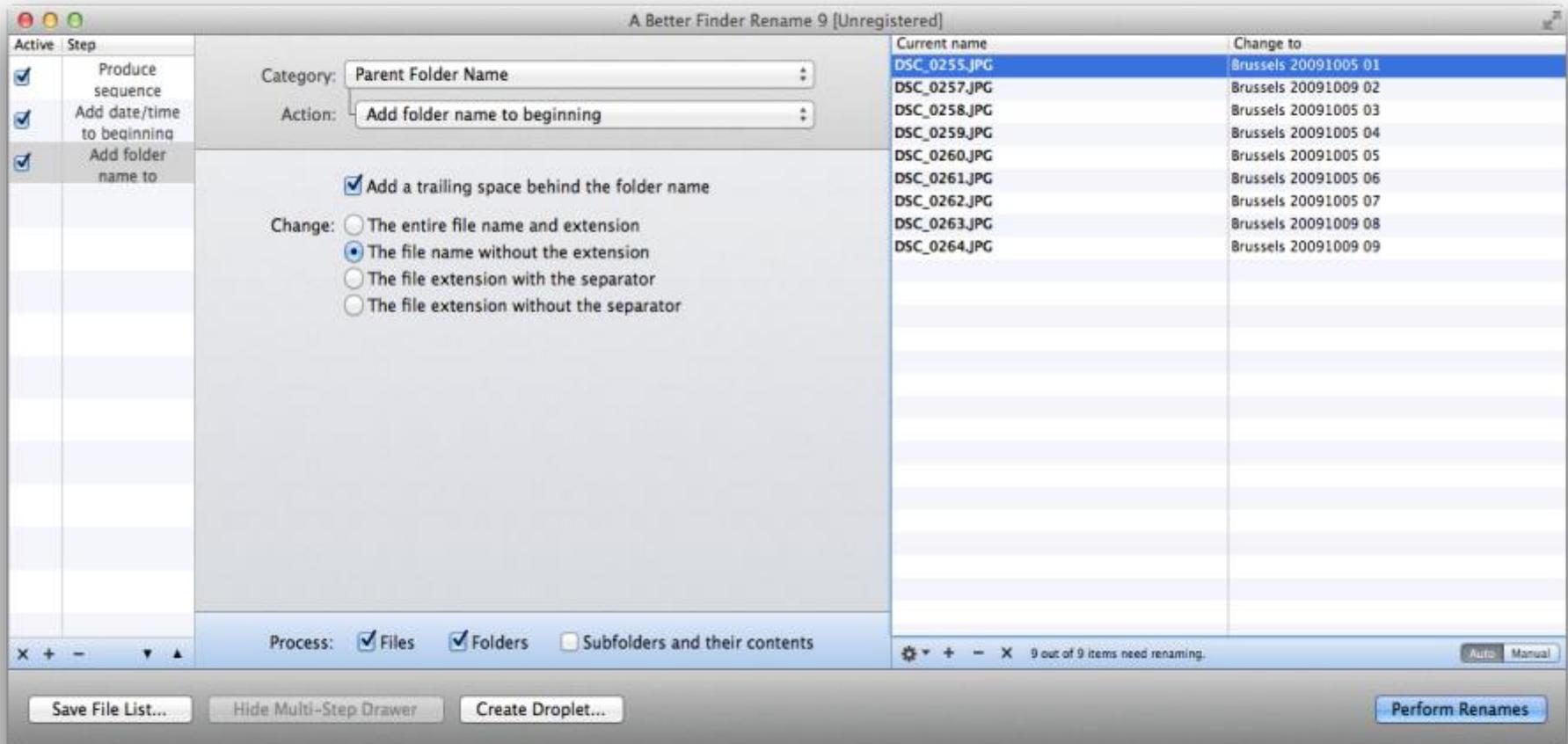
# Basic window: Adding a sequence



# Verification



# Multi-Step Feature

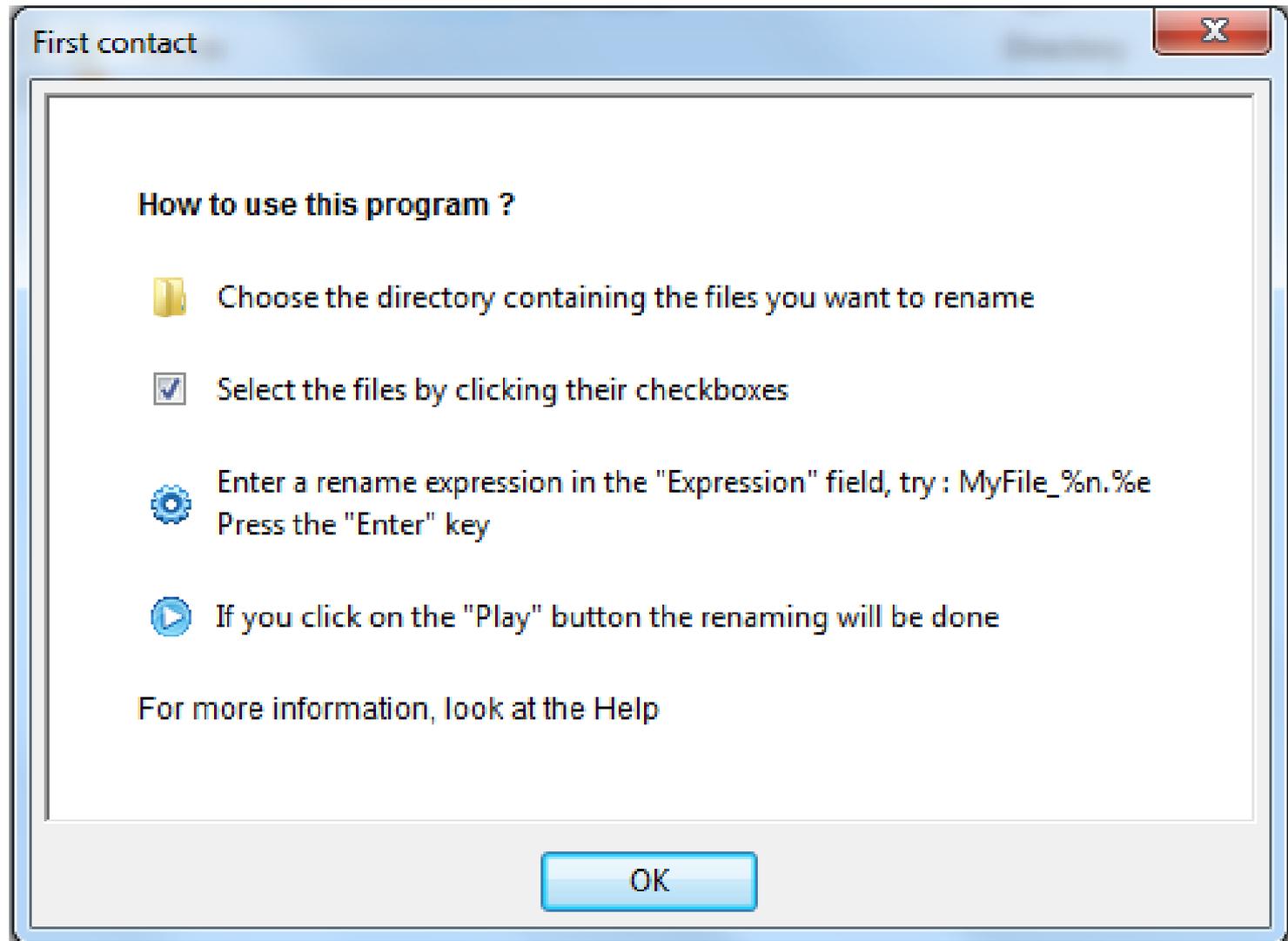


# Siren

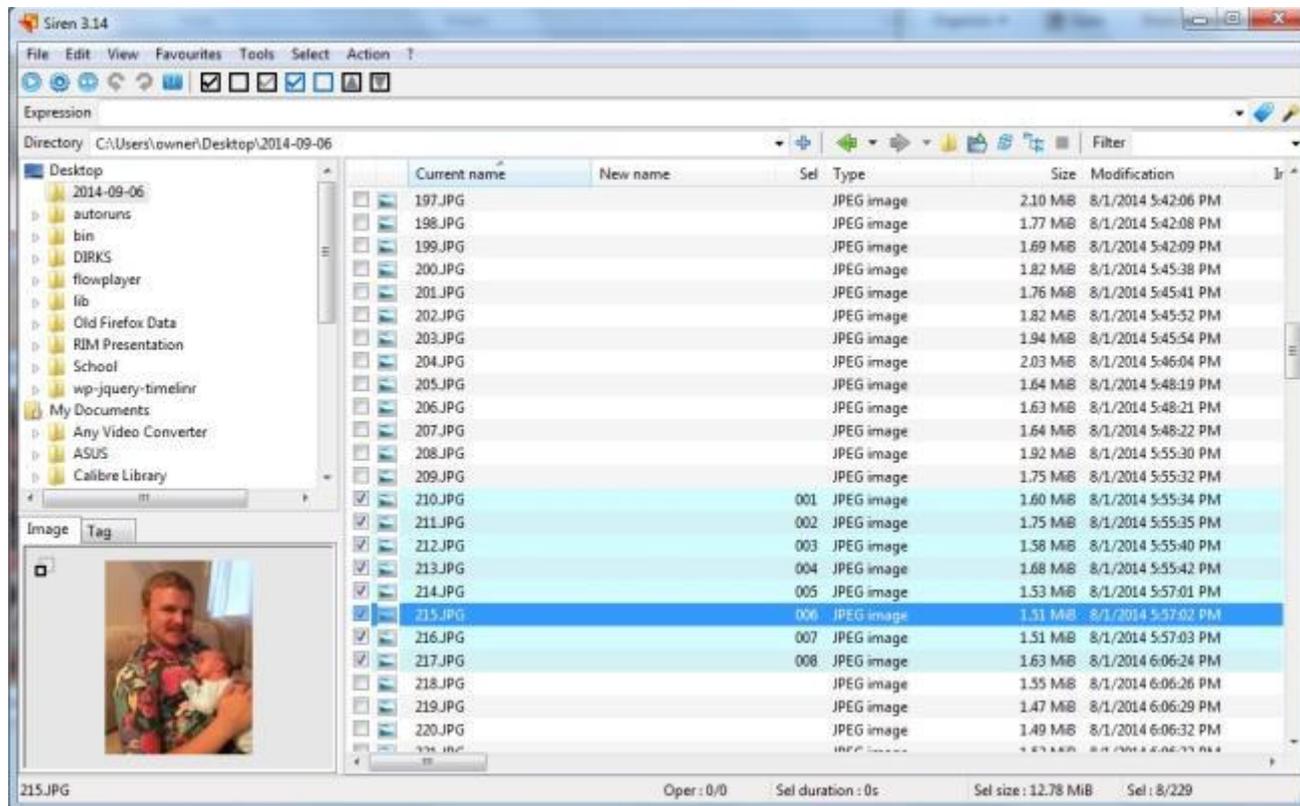
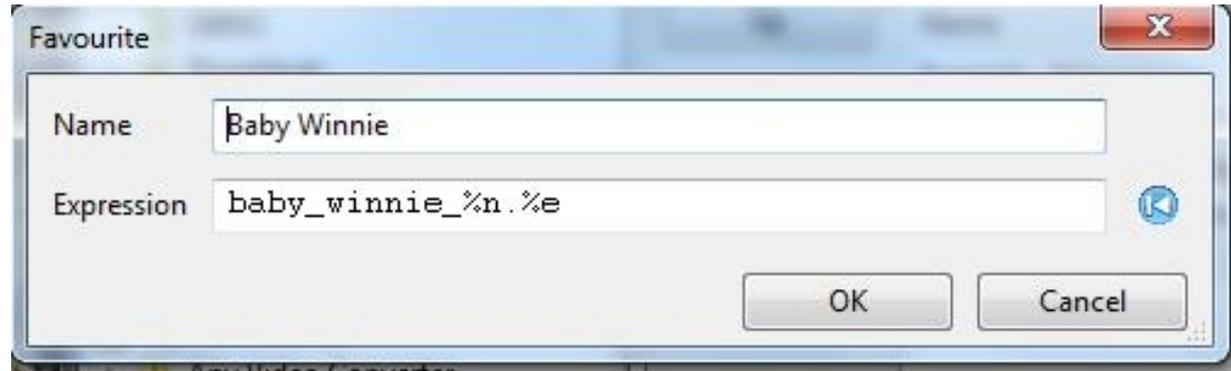


- Operating System: Windows
- Developer: Scarabee-Software (France)
- Price: Free
- Features:
  - Rename, move and copy files to other locations
  - Windows Explorer based folder structure
  - Customizable metadata based renaming
    - Ex 1: Music files can be renamed according to ID3 tags
    - Ex 2: Picture files can be renamed according to EXIF info.
- Limitations:
  - Weak documentation
- Download: <http://www.softpedia.com/get/Office-tools/Other-Office-Tools/Siren.shtml>

# First Contact Pop Up



# Interface



# After customized renaming



baby\_winnie\_002



baby\_winnie\_003



baby\_winnie\_004



baby\_winnie\_005



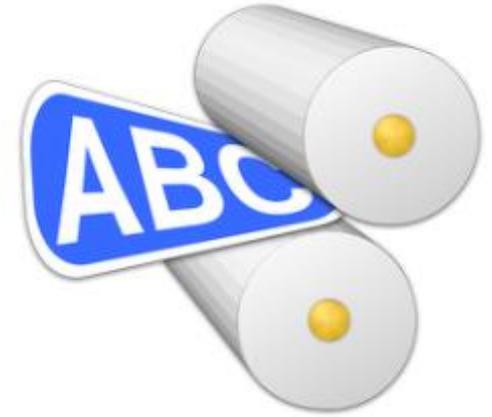
baby\_winnie\_006



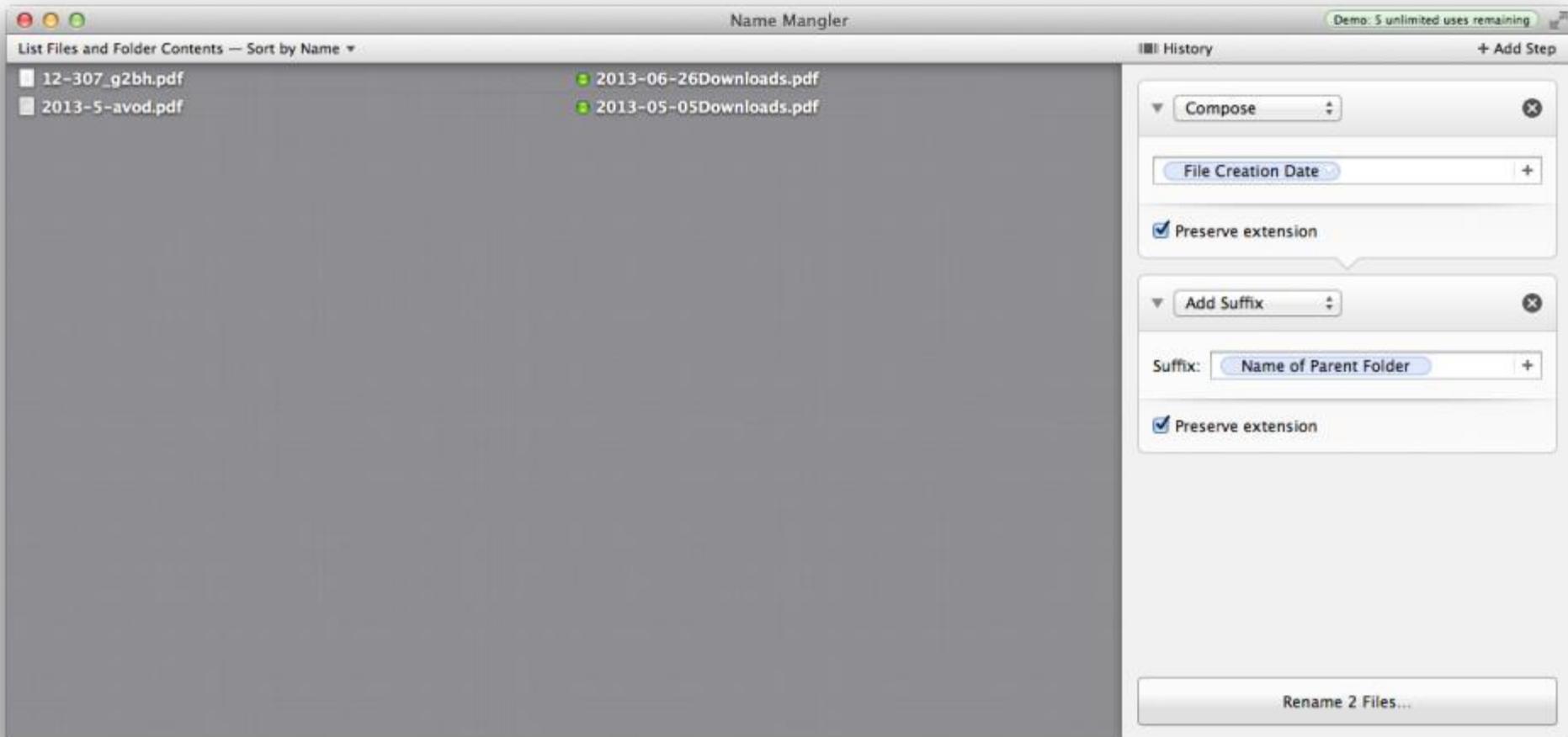
baby\_winnie\_007

# Name Mangler

- Operating System: Mac
- Developer: Many Tricks
- Price: Free trial, \$19 per license
- Features:
  - Multi-step renaming
  - Use over 150 metadata fields
  - Presets and droplets
  - Preview all file name changes
- Limitations:
  - Trial only allows 5 uses before purchase
  - Small thumbnails
- Download: <http://manytricks.com/namemangler/>



# Basic window



# Renaming Options

The screenshot shows the 'Name Mangler' application window. The main area displays a list of files and folders with their names and creation dates. The right-hand panel contains settings for renaming, including a 'Sequence' dropdown, 'Format' (Arabic: 1, 2, 3, ...), 'Start at' (1), 'Increase by' (1), 'Number of digits' (Auto), 'Prefix' (File Creation Date), and 'Suffix'. A 'Preserve extension' checkbox is checked. A 'Rename 10 Files...' button is at the bottom right.

File Name	Creation Date
IMG_2027.JPG	2014-01-02_01.JPG
IMG_2029.JPG	2014-01-02_02.JPG
IMG_2030.JPG	2014-01-02_03.JPG
IMG_2031.JPG	2014-01-02_04.JPG
IMG_2032.JPG	2014-01-02_05.JPG
IMG_2033.JPG	2014-01-02_06.JPG
IMG_2034.JPG	2014-01-02_07.JPG
IMG_2036.JPG	2014-01-03_08.JPG
IMG_2037.JPG	2014-01-03_09.JPG
IMG_2038.JPG	2013-12-24_10.JPG

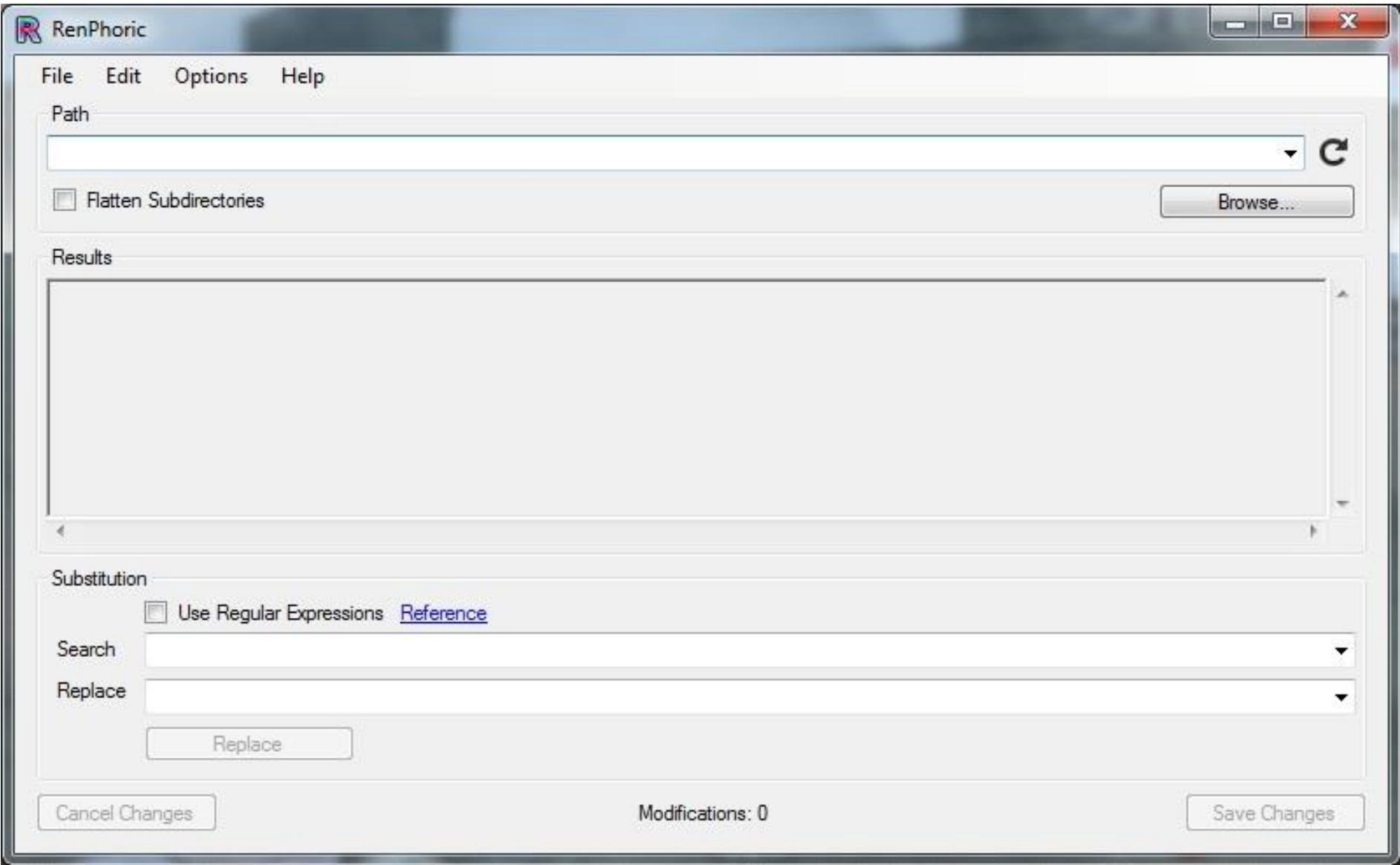
- ✓ Find and Replace
- Sequence
- Compose
- Add Prefix
- Add Suffix
- Insert
- Remove**
- Change Case
- Set Extension
- Advanced

# RenPhoric

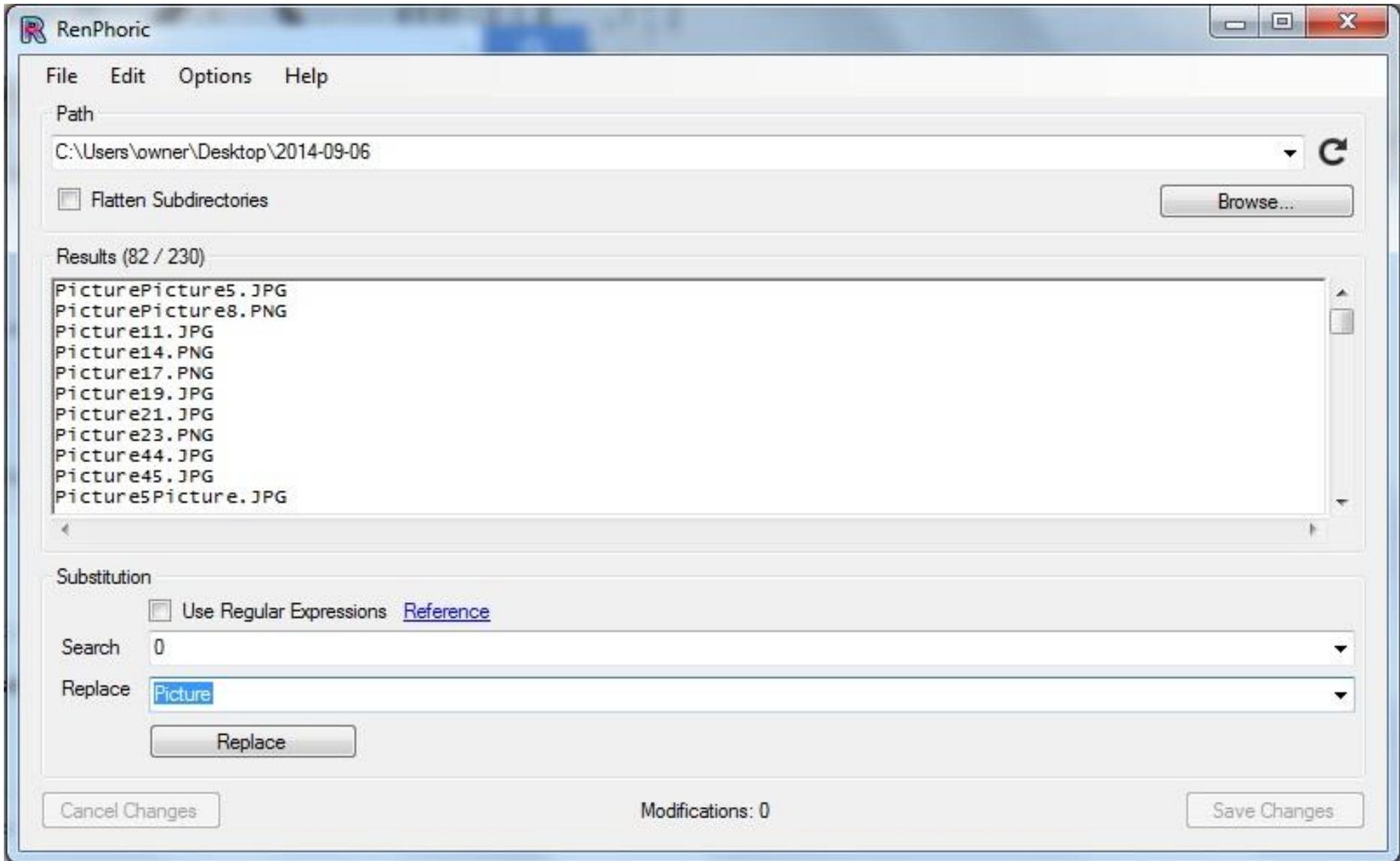


- Operating System: Windows
- Developer: neophoric
- Price: Freeware
- Features:
  - Find and Replace parts of file names.
  - Quick and easy to use.
- Limitations:
  - Can only search and replace parts of file names.
- Download: <http://www.neophoric.com/renphoric/>

# RenPhoric Layout



# Open a directory to search and replace



# Bad use of Search and Replace



2Picture3



2Picture4



2Picture5



2Picture6



2Picture7



2Picture8

# Other bad uses



baby\_winnie\_PicturePicture1



baby\_winnie\_PicturePicture2



baby\_winnie\_PicturePicture3



baby\_winnie\_PicturePicture4



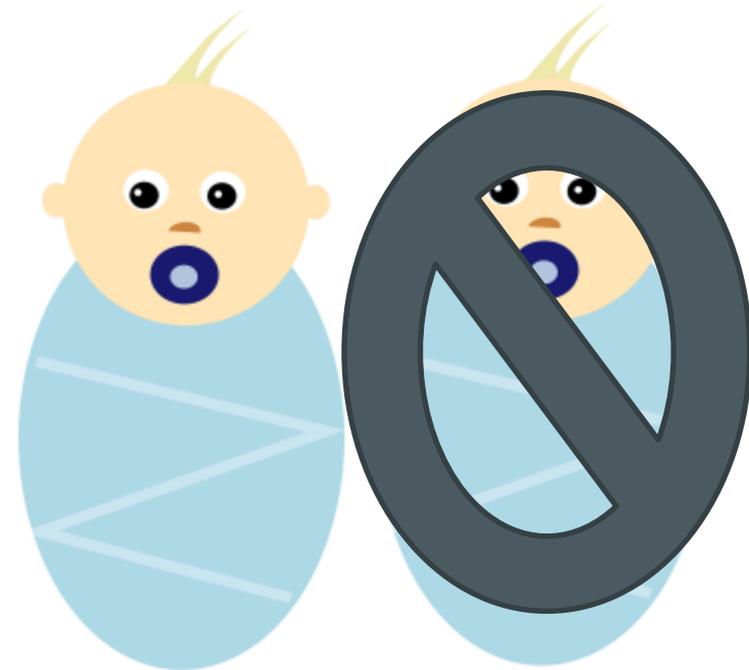
baby\_winnie\_PicturePicture5



baby\_winnie\_PicturePicture6

# De-duplicating

1. Anti Twin (Windows)
2. dupeGuru (Mac, Windows, Linux, Ubuntu)
3. Duplicate Cleaner (Windows)
4. PhotoSweeper (Mac)
5. Duplicate File Finder (Windows)
6. MacBooster 2 (Mac)
7. NoDupe (Windows)



# Anti-Twin



- Operating System: Windows
- Developer: Jorg Rosenthal Software
- Price: Free for Private Use
- Features:
  - Search by file name, byte by byte or pixel by pixel (for pictures)
  - Option to review and then delete files
  - Helpful pop-up tip sheet
- Limitations:
  - Slowest Windows tool. Uses bubble sort comparison.
  - Questionable results when using pixel by pixel comparison.
- Download: <http://www.joerg-rosenthal.com/en/antitwin/>

# Layout

Tips

## Tips

### Big fish first

If you want results as fast as possible, let Anti-Twin compare only large files first. To do this, set the minimum file size to e.g. 500 kb in the "File filter".

### Set your target

As the basic folder, do not select the whole hard disk C:\. Instead, specifically select the relevant folder(s). Otherwise you might find the long search times and the complex results frustrating.

### Hands off the system folders

Although it may be tempting, but please never delete files from the Windows or Program folders! Only delete files that you created, e.g. pictures, music or documents.

### MP3s : less is more

When comparing MP3-files, Anti-Twin can strategically ignore ID tags containing information about the title and the artist. To activate this, set the match rate for the content comparison to 95%.

Find more information online:

[FAQ](#) | [User Guide](#) | [Technical Details](#)

Anti-Twin v 1.8d Find and delete duplicate files (C) Jörg Rosenthal, 2003 - 2011

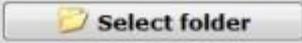
File View Control Language About

Start search Result Licensed for: Private user

Here please select in which folders you wish to search for files and the comparison type to be applied [Website](#)

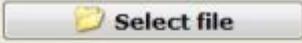
**Which files do you wish to compare?** [Help](#)

Basic folder C:\Users\owner\Documents 

2nd folder 

Include subfolders  Do not filter, i.e. check all files

Compare all files  Only files in different (sub)folders  Only files in different basic folders

Search for a certain file 

**Which comparison method(s)?** [Help](#)

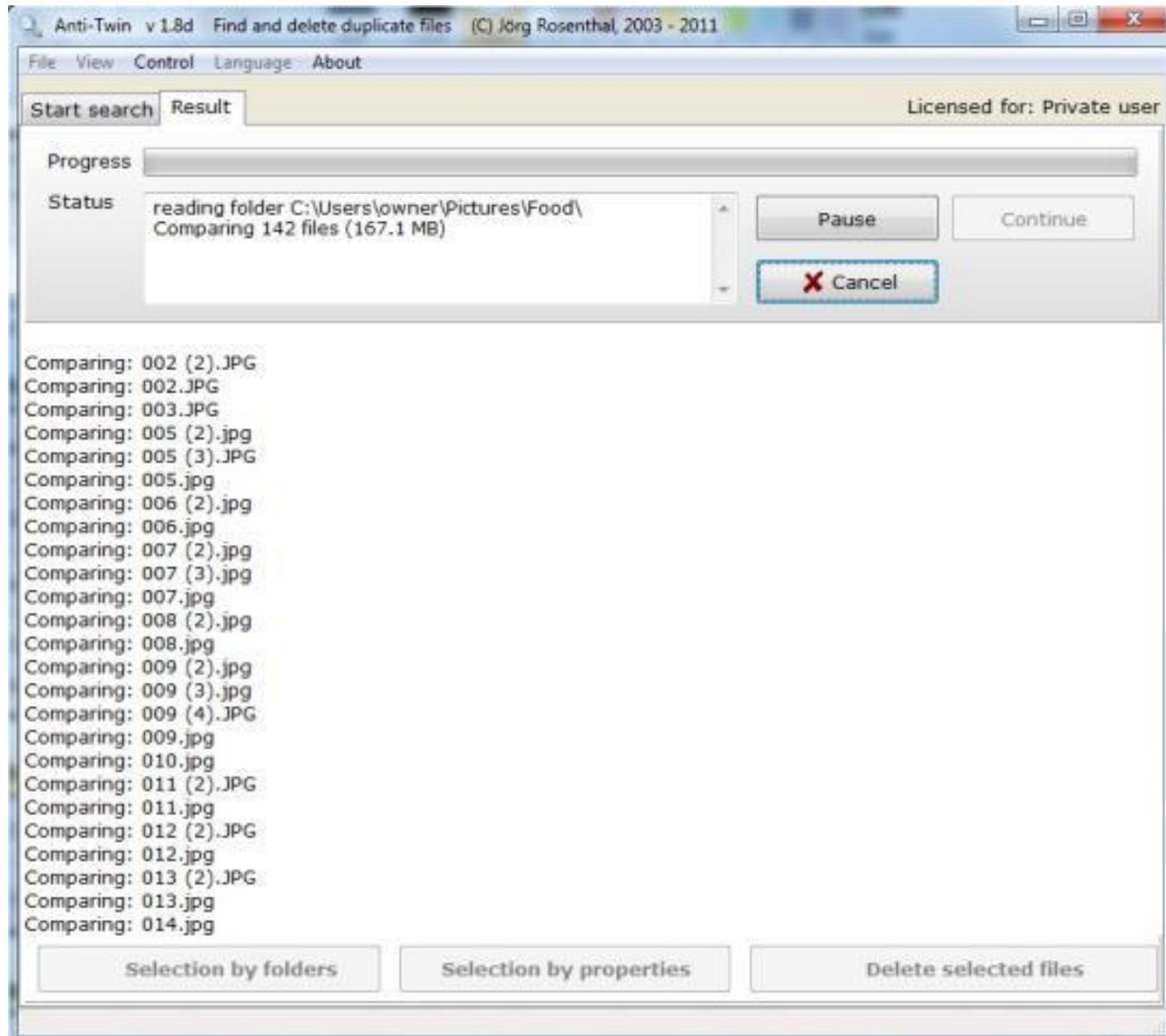
Compare names  Different extensions  Identical file size Match min. 100%

Compare content  Byte by Byte  Compare images (pixels) Match min. 100%

**Start Anti-Twin ...**

Limit to 1000 duplicates  

# Bubble Sort Visual



# Pixel by Pixel Results

Anti-Twin v 1.8d Find and delete duplicate files (C) Jörg Rosenthal, 2003 - 2011

File View Control Language About

Start search Result

Licensed for: Private user

Progress

Status reading folder C:\Users\owner\... Comparing 142 files (167.1%) 40 similar files found (25.5%)

■ > 007 (2).jpg	2013-06-18	459 KB
■ 95% 020.jpg	2013-07-26	30 KB
■ 99% 368.jpg	2013-06-18	1,781 KB
■ > 018 (2).jpg	2011-03-24	1,629 KB
■ 99% Food.jpg	2011-03-23	56 KB
■ > 111.jpg	2012-07-19	2,299 KB
■ 94% 320.jpg	2012-07-19	37 KB
■ 99% 321.jpg	2012-07-19	2,299 KB
■ > 112.jpg	2012-07-20	1,474 KB
■ 92% 334.jpg	2012-07-20	36 KB
■ 99% 335.jpg	2012-07-20	1,474 KB
■ > 113.jpg	2012-07-21	2,076 KB
■ 93% 338.jpg	2012-07-21	40 KB
■ 99% 339.jpg	2012-07-21	2,076 KB
■ > 176.jpg	2012-08-12	1,848 KB
■ 95% 490.jpg	2012-08-12	36 KB
■ 99% 491.jpg	2012-08-12	1,848 KB
■ > 183.jpg	2012-08-17	1,612 KB
■ 95% 507.jpg	2012-08-17	33 KB
■ 99% 508.jpg	2012-08-17	1,612 KB
■ > 203.jpg	2012-09-15	2,076 KB

Image Preview

Hide Disable



176.jpg  
C:\Users\owner\Pictures\Food\  
3264 x 2448 pixel, 1847 kb



490.jpg  
C:\Users\owner\Pictures\Food\  
408 x 306 pixel, 36 kb

Pause Continue

Cancel

Selection by folders Selection by properties Delete selected files

Tips

**Number One Choice**

From the list of results, select the redundant files you wish to delete or move. Select the checkbox in front of the relevant files to mark them for deletion (red).

Alternatively you can use the function "Selection by folders". There please define the folders (and subfolders) in which you want Anti-Twin to delete duplicate files.

Find more information online:  
[Select for Deletion](#)

Legend

- File will not be removed (default setting) but selection can be changed to a "delete" selection with the function »Selection by ...«
- File will not be removed (manually set)
- ⊗ Delete/move file
- ⊗ File has already been deleted

# Same folder using byte by byte

The screenshot displays a file comparison application window with a menu bar (File, View, Control, Language, About) and a license notice (Licensed for: Private user). The main window is divided into 'Start search' and 'Result' tabs. The 'Result' tab shows a progress bar and status information: 'reading folder C:\Users\owner\Pictures\Food\ Comparing 142 files (167.1 MB) 1 similar files found (2.4 MB)'. Below this is a table of results:

File Name	Date	Size
> 223.jpg	2012-10-17	2,476 KB
94% 633.jpg	2012-10-17	2,476 KB

A 'Legend' dialog box is open in the foreground, providing a key for the file status indicators:

- File will not be removed (default setting) but selection can be changed to a "delete" selection with the function »Selection by ...«
- File will not be removed (manually set)
- Delete/move file
- File has already been deleted

At the bottom of the application window, there are three buttons: 'Selection by folders', 'Selection by properties', and 'Delete selected files'.

## Tips

### Number One Choice

From the list of results, select the redundant files you wish to delete or move. Select the checkbox in front of the relevant files to mark them for deletion (red).

Alternatively you can use the function "Selection by folders". There please define the folders (and subfolders) in which you want Anti-Twin to delete duplicate files.

Find more information online:

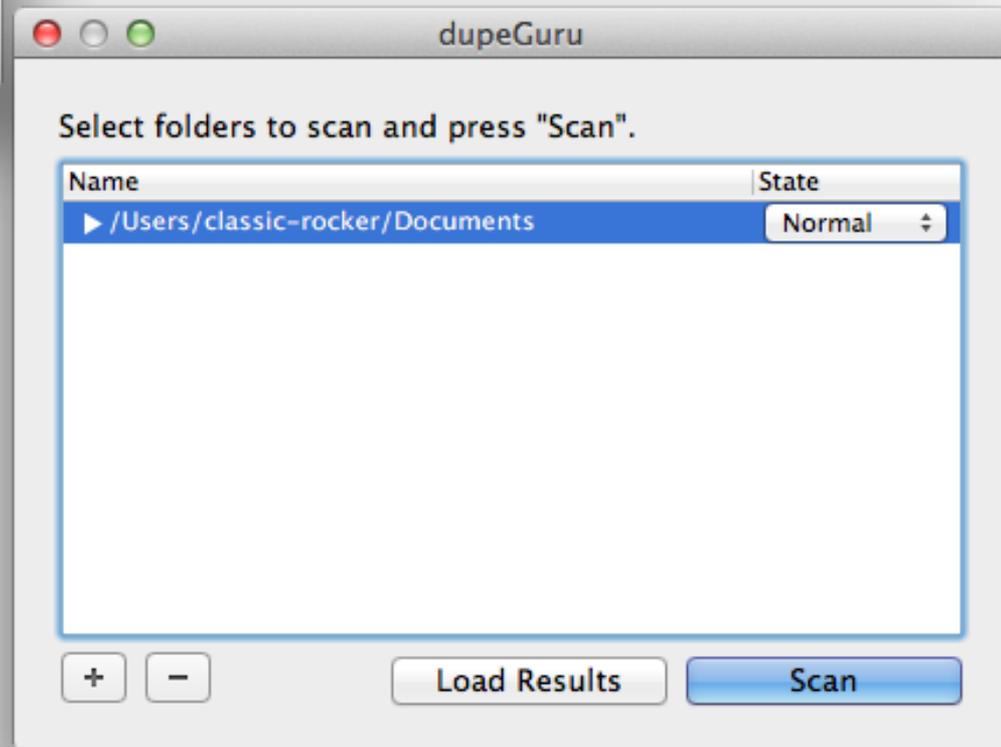
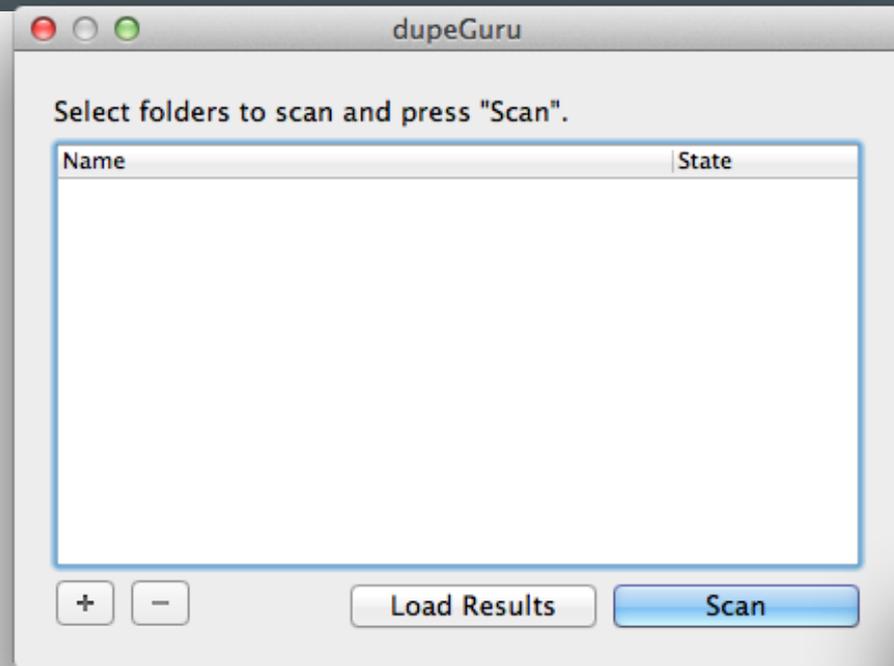
[Select for Deletion](#)

# dupeGuru



- Operating System: Mac, Ubuntu, Arch Linux, Windows
- Developer: Hardcoded Software
- Price: Free (Shareware/Open Source)
- Features:
  - Can scan file names or contents
  - “Fuzzy matching algorithm” for highly similar file names
- Limitations:
  - Slow
  - Less visually pleasing interface
  - No thumbnails
- Download: <http://www.hardcoded.net/dupeguru/>

# Basic window



Scan multiple folders at once



# Scan results

dupeGuru Results

Details Dupes Only Delta Filter

Filename	Folder	Size (KB)	Match %
<input type="checkbox"/> Sies and Silver.doc	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	33	100
<input type="checkbox"/> Sies and Silver.doc	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	33	100
<input type="checkbox"/> ~WRL2311.tmp	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	33	100
<input type="checkbox"/> ~WRL2311.tmp	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	33	100
<input type="checkbox"/> City Planning and Infrastructure.pdf	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	241	100
<input type="checkbox"/> City Planning and Infrastructure.pdf	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	241	100
<input type="checkbox"/> DSC_0262.JPG	/Users/classic-rocker/Documents/Grad School/_Classes/3_2014_Fall/INF3...	4289	100
<input type="checkbox"/> Brussels 20091005 07.JPG	/Users/classic-rocker/Documents/Grad School/_Classes/3_2014_Fall/INF3...	4289	100
<input type="checkbox"/> Oracle_Bone_Questions-Shang_D...	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	69	100
<input type="checkbox"/> Oracle_Bone_Questions-Shang_D...	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	69	100
<input type="checkbox"/> KP65-805c.jpg	/Users/classic-rocker/Documents/Job Materials/2014_Pixar/PhotoLibrary/t...	4825	100
<input type="checkbox"/> KP65-805c.jpg	/Users/classic-rocker/Documents/Job Materials/2014_Pixar/PhotoLibrary/t...	4825	100
<input type="checkbox"/> ~WRL4018.tmp	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	59	100
<input type="checkbox"/> ~WRL4018.tmp	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	59	100
<input type="checkbox"/> Modern Industrial City Brochure...	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	4069	100
<input type="checkbox"/> Modern Industrial City Brochure...	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	4069	100
<input type="checkbox"/> Primary Source Analysis 3.doc	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	46	100
<input type="checkbox"/> Primary Source Analysis 3.doc	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	46	100
<input checked="" type="checkbox"/> 1904 Gramercy 3.JPG	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	2449	100
<input type="checkbox"/> 1904 Gramercy 3.JPG	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	2449	100
<input type="checkbox"/> Annotated Bibliography.doc	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	118	100
<input type="checkbox"/> Annotated Bibliography.doc	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	118	100
<input type="checkbox"/> KP65-808c.jpg	/Users/classic-rocker/Documents/Job Materials/2014_Pixar/PhotoLibrary/t...	2873	100
<input type="checkbox"/> KP65-808c.jpg	/Users/classic-rocker/Documents/Job Materials/2014_Pixar/PhotoLibrary/t...	2873	100
<input type="checkbox"/> 8-1912 Craftsman page 470.jpg	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	70	100
<input type="checkbox"/> 8-1912 Craftsman page 470.jpg	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	70	100
<input type="checkbox"/> LA Times 10-27-1912.pdf	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	129	100
<input type="checkbox"/> LA Times 10-27-1912.pdf	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	129	100
<input type="checkbox"/> 8-1912 Craftsman page 472.jpg	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	113	100
<input type="checkbox"/> 8-1912 Craftsman page 472.jpg	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	113	100
<input type="checkbox"/> LA Times 2-23-1913.pdf	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	49	100
<input type="checkbox"/> LA Times 2-23-1913.pdf	/Users/classic-rocker/Documents/Westmont 2008-2012/2010-2011 Third...	49	100

0 / 235 (0.00 B / 194.44 MB) duplicates marked.

# Options

Send Marked to Trash...

Move Marked to...

Copy Marked to...

Remove Marked from Results

Remove Selected from Results

Add Selected to Ignore List

Make Selected into Reference

Open Selected with Default Application

Reveal Selected in Finder

Rename Selected

# Duplicate Cleaner - Free



- Operating System:
- Developer: Digital Volcano (U.K.)
- Price: Free with limited features. \$29.99 for full features
- Features:
  - Best looking interface
  - Accurate
  - Fast
- Limitations:
  - Can't do pixel by pixel comparison without paying
  - Preview of pictures not as useful as others
  - A lot of clicking required
- Download:

<http://www.digitalvolcano.co.uk/dcinstalled.html>

# Layout

The screenshot displays the Duplicate Cleaner Free 3.2.4 application window. The title bar shows the application name and standard Windows window controls. The menu bar includes File, View, Options, and Help. Below the menu bar is a toolbar with icons for About, Scan Now, Image Preview, File Removal..., Upgrade to Pro, More Options..., and Exit. The version number 3.2.4 is displayed in the top right corner.

The main interface is divided into several sections:

- Search Criteria:** A panel with a monitor icon and the text "What to find".
- Scan Location:** A panel with a folder icon and the text "Where to look".
- Duplicate Files:** A panel with a document icon and the text "Nothing Scanned Yet".
- Duplicate Folders:** A panel with a folder icon and the text "Nothing Scanned Yet".

Below these panels are three tabs: Regular Mode (selected), Image Mode, and Audio Mode. The Regular Mode panel contains two sections:

- Find Files With:** Radio buttons for "Same Content" (selected) and "Ignore Content".
- Additional Options:** Checkboxes for "Same File Name", "Similar File Names", "Same Size", "Same Created Date", and "Same Modified Date".

The "More Search Options" panel (indicated by a gear icon) contains the following checkboxes:

- Exclude Zero Size Files from Duplicate List
- Don't Scan System Files/Folders
- Don't Follow NTFS Mountpoints and Junctions
- Count Hardlinks in File
- Exclude Hard Linked files from Duplicate List

The right side of the interface features a large banner for "DuplicateCleaner Free" with a "Non-commercial use only" logo. Below the banner is the "Search Filters" section, which includes:

- File Filter:** "Included:" dropdown set to "\*\*" and "Excluded:" dropdown.
- File Sizes:** "Any Size" checked, with "Minimum File Size (KB)" and "Maximum File Size (KB)" spinners both set to 0.
- File Dates:** "Any Date" selected, with "Between" and "And" date pickers both set to 10/11/2014.

# Finding Duplicates

The image shows the Duplicate Cleaner Free 3.2.4 software interface. The main window displays the results of a scan, indicating that 104/104 files were scanned (97.8 MB), 1 group of duplicates was found, and 2 files have duplicates (136 KB). The total time taken for the scan was 00:00:01. The software is currently showing 40 duplicate files, with 20 files marked for deletion, totaling 628 KB. The interface includes a menu bar (File, View, Options, Help) and a toolbar with buttons for About, Scan Now, Image Preview, File Removal, Upgrade to Pro, More Options, and Exit. The main area is divided into sections for Search Criteria, Scan Location, Duplicate Files, and Duplicate Folders. A Selection Assistant is active, showing a list of files with columns for File Name, Path, Group, Size, Created, and Modified. The status bar at the bottom indicates 'Ready. 1 File(s) Selected 20 File(s) Marked'.

**Scan Complete**

Scan Details

- 104/104 Files Scanned (97.8 MB)
- 1 Groups of duplicates
- 2 Files have duplicates(136 KB)
- Total Time Taken: 00:00:01

**Duplicate Cleaner Free 3.2.4**

File View Options Help

About Scan Now Image Preview File Removal... Upgrade to Pro More Options... Exit

Version 3.2.4

Search Criteria What to find

Scan Location Where to look

Duplicate Files 40 Duplicate Files

Duplicate Folders

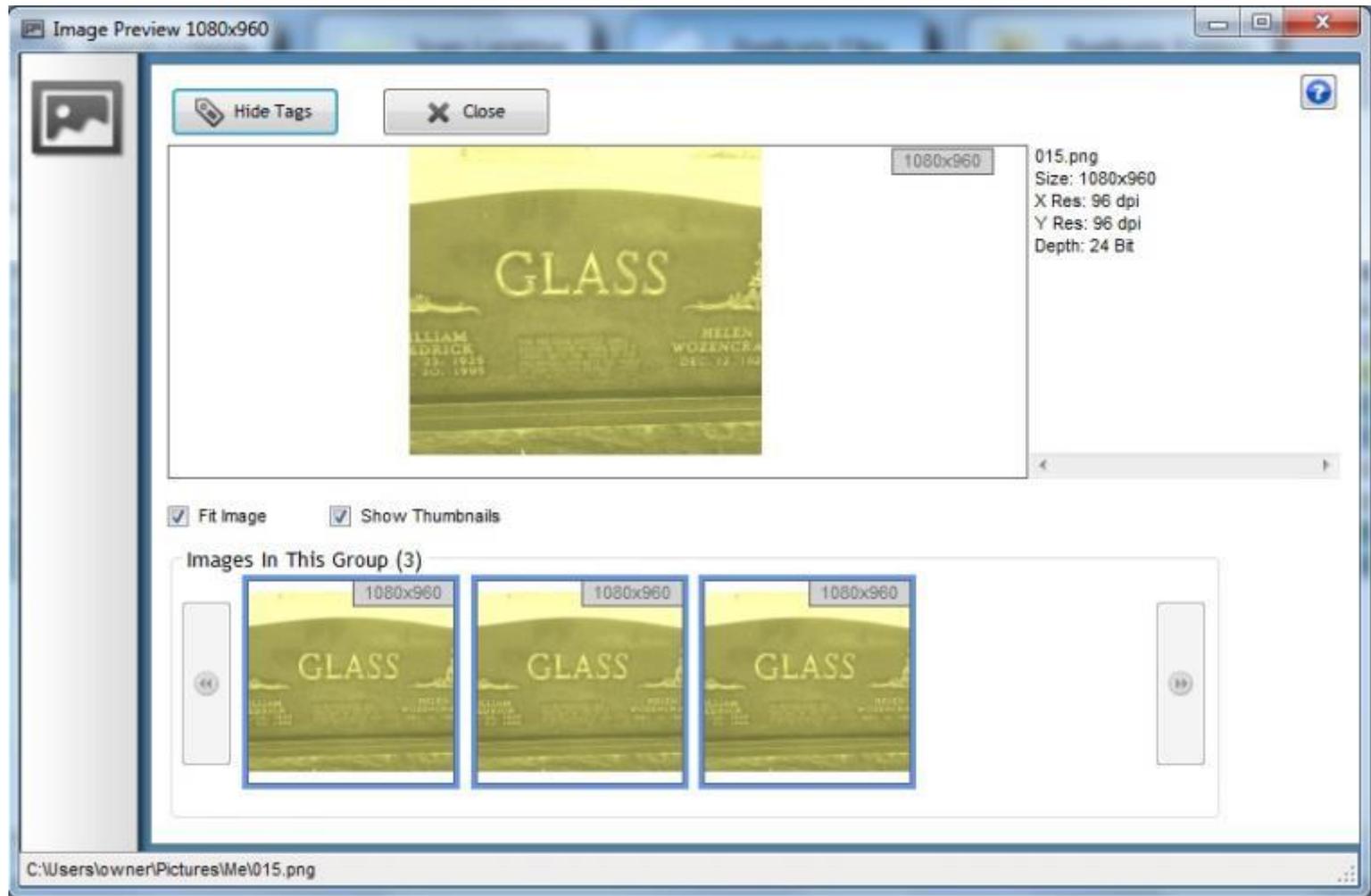
Show Summary Selection Assistant Quick Columns

20 files marked for deletion, total size 628 KB Filter Marked: All Filter Other: No Filters

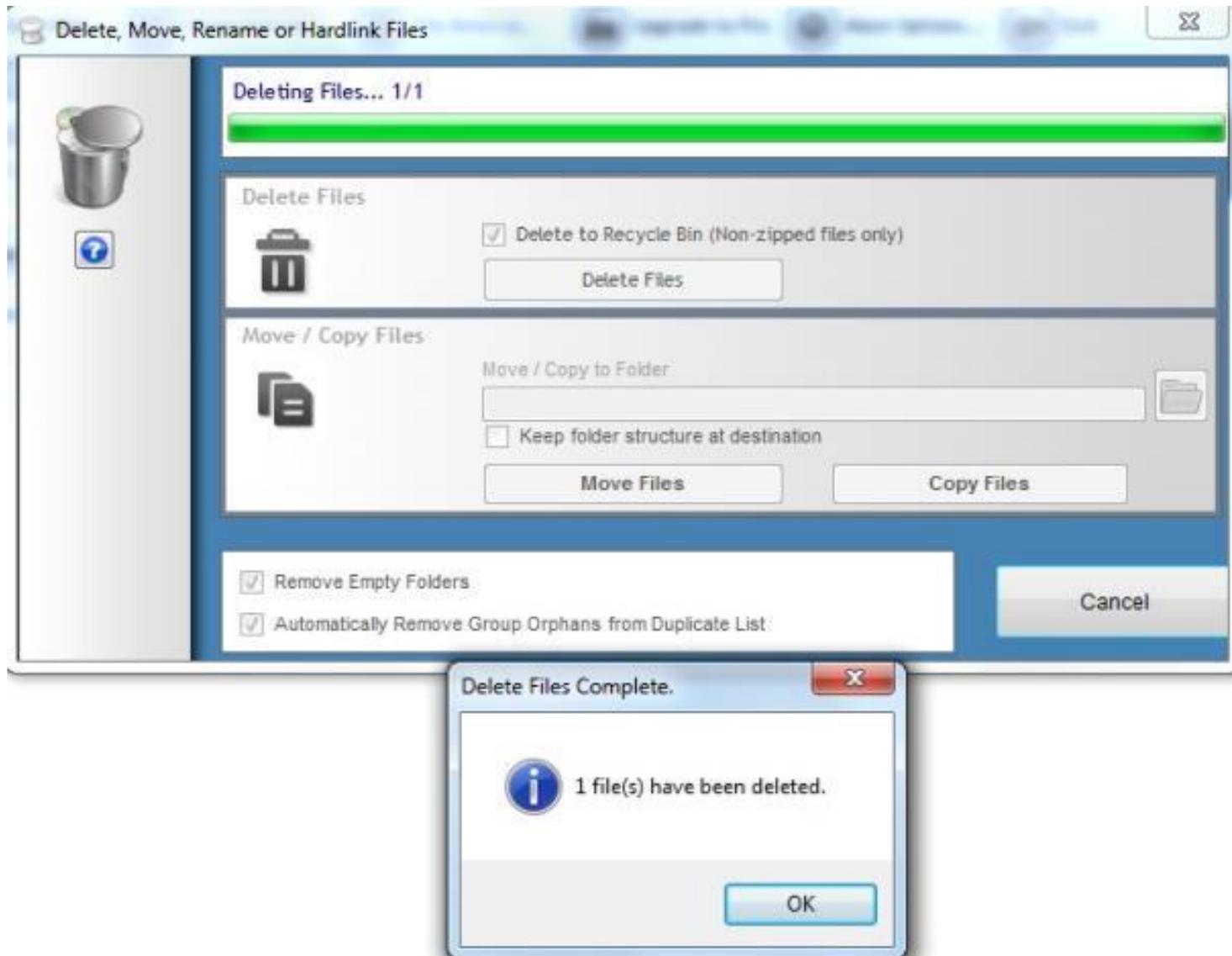
File Name	Path	Group	Size	Created	Modified
511.jpg	C:\Users\owner\Pictures\Environme...	13	34.2 KB	8/23/2012 3:20:36	8/23/2012 3
517.jpg	C:\Users\owner\Pictures\Environme...	13	34.2 KB	8/23/2012 4:28:28	8/23/2012 4
560.jpg	C:\Users\owner\Pictures\Environme...	14	26.5 KB	9/10/2012 3:10:06	9/10/2012 3
564.jpg	C:\Users\owner\Pictures\Environme...	14	26.5 KB	9/11/2012 7:56:58	9/11/2012 7
607.jpg	C:\Users\owner\Pictures\Environme...	15	45.0 KB	10/6/2012 4:44:24	10/6/2012 4
609.jpg	C:\Users\owner\Pictures\Environme...	15	45.0 KB	10/6/2012 4:44:28	10/6/2012 4
616.jpg	C:\Users\owner\Pictures\Environme...	16	34.2 KB	10/8/2012 7:56:02	10/8/2012 7
618.jpg	C:\Users\owner\Pictures\Environme...	16	34.2 KB	10/8/2012 7:56:08	10/8/2012 7
626.jpg	C:\Users\owner\Pictures\Environme...	17	32.5 KB	10/8/2012 8:34:24	10/8/2012 8
628.jpg	C:\Users\owner\Pictures\Environme...	17	32.5 KB	10/8/2012 8:34:28	10/8/2012 8
669.jpg	C:\Users\owner\Pictures\Environme...	18	35.1 KB	10/20/2012 8:15:34	10/20/2012
671.jpg	C:\Users\owner\Pictures\Environme...	18	35.1 KB	10/20/2012 8:15:44	10/20/2012
674.jpg	C:\Users\owner\Pictures\Environme...	19	25.0 KB	10/20/2012 8:43:56	10/20/2012
676.jpg	C:\Users\owner\Pictures\Environme...	19	25.0 KB	10/20/2012 8:44:02	10/20/2012
689.jpg	C:\Users\owner\Pictures\Environme...	20	34.5 KB	10/22/2012 6:50:00	10/22/2012
691.jpg	C:\Users\owner\Pictures\Environme...	20	34.5 KB	10/22/2012 6:50:16	10/22/2012

Ready. 1 File(s) Selected 20 File(s) Marked

# Duplicate Preview Screen



# Deleting a Duplicate



# PhotoSweeper

- Operating System: Mac
- Developer: Overmacs
- Price: Free demo, \$9.99 for full version
- Features:
  - Integrates with iPhoto, Adobe Lightroom, Aperture
  - Finds identical and highly similar photos
  - Comparison and Mark for Deletion options
- Limitations:
  - Only for photos, not for other files
  - Demo limits you to 101 photos at a time
- Download:

<http://overmacs.com/photosweeper.html>

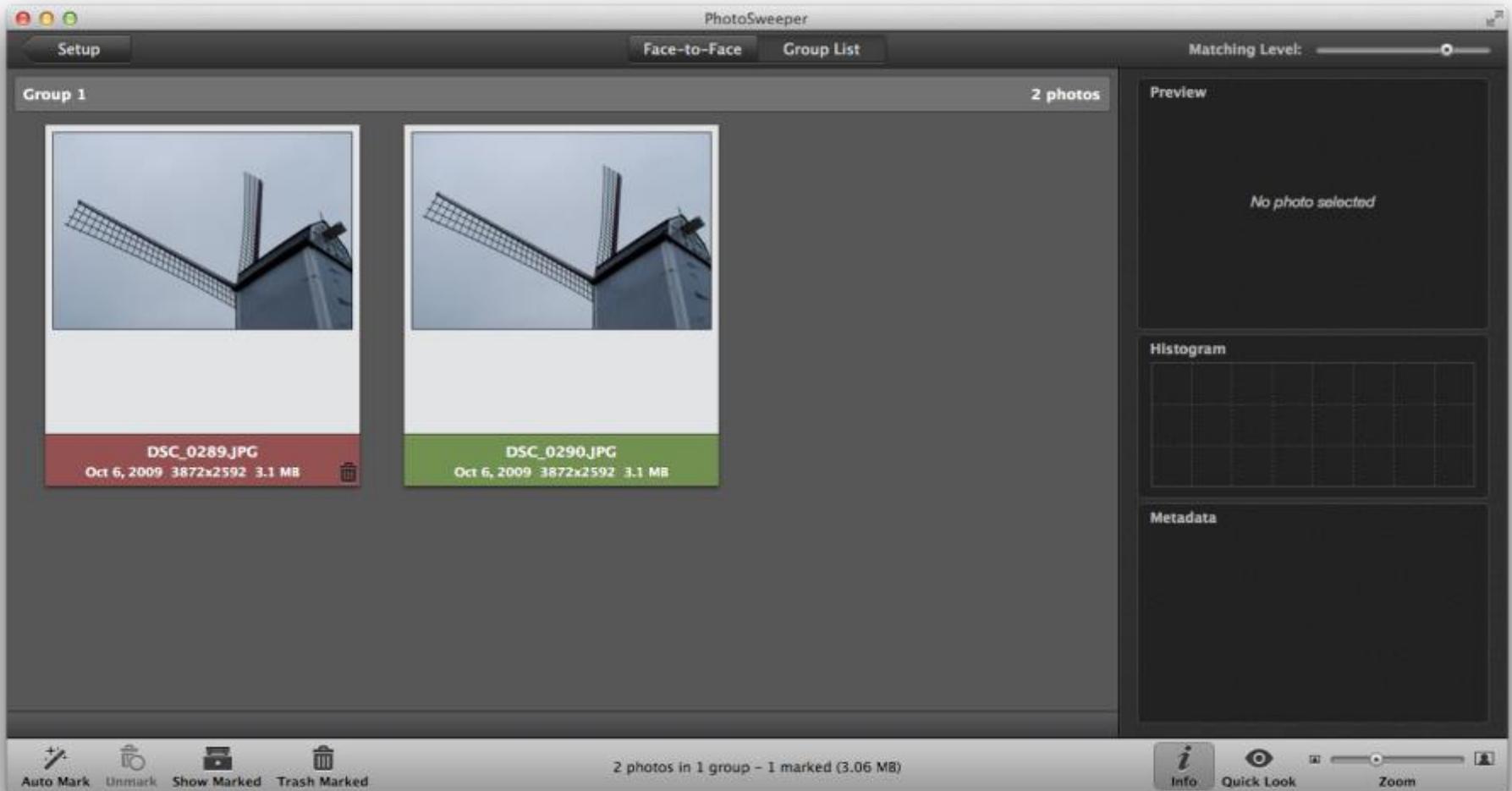


# Main window

- Bitmap
- Histogram
- Time Interval
- Time + Bitmap
- ✓ Time + Histogram
- Duplicates Only

The screenshot displays the PhotoSweeper application window. The title bar reads "PhotoSweeper". The status bar at the top left indicates "87 photos - 337.30 MB on disk". The main area contains a grid of photo thumbnails. The first row shows three photos: a blue door (DSC\_0281.JPG), a close-up of the same door (DSC\_0283.JPG), and a street scene (DSC\_0284.JPG). The second row shows a smaller version of the street scene, a close-up of a windmill, and another view of the windmill. The right sidebar has two tabs: "Settings" and "Image Info". The "Settings" tab is active, showing a "Comparison Method" dropdown set to "Time + Bitmap" with a hint: "Hint: Performs 'Time Interval' method at first, and then 'Bitmap' method for received results." Below this are sliders for "Time Interval" (set to "1 minute and less") and "Bitmap Size" (set to "48x48 pixels"). The "Details" section includes "Color Mode" (set to "RGB"), "Preprocessing" (set to "Align Colors + Blur"), and "Color Sensitivity" (set to "Average"). At the bottom of the sidebar is a "Matching Level" slider set to "Exact matching". The bottom of the window features a "Media Browser" icon, an "Add Photos" button, a "Compare" button, and icons for "Info", "Quick Look", and "Zoom".

# Finding highly similar photos

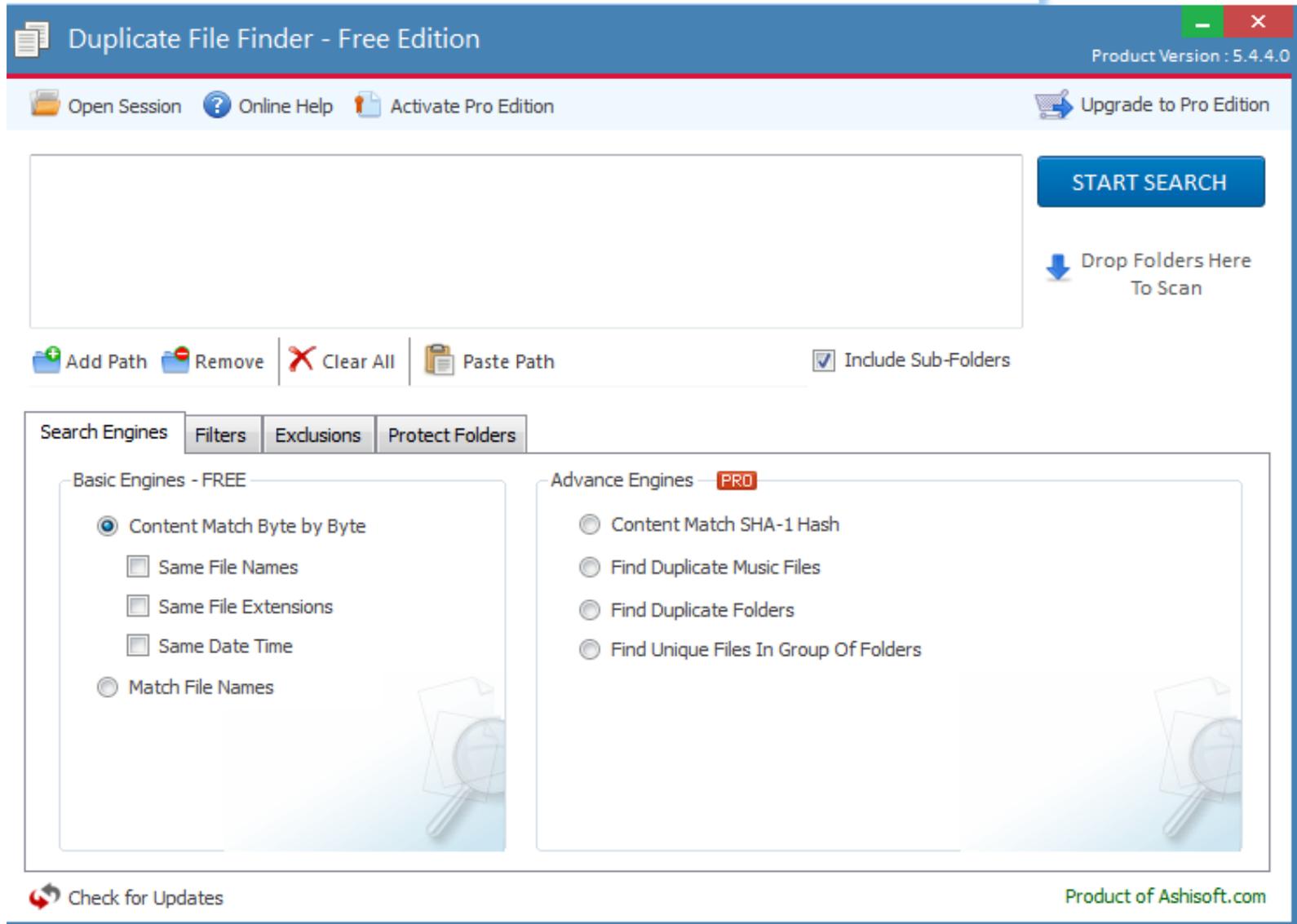


# Duplicate File Finder - Free



- Operating System: Windows
- Developer: Ashisoft
- Price: Free with limited features. \$ 29.99 for full functionality.
- Features:
  - Analyzes files byte by byte to find duplicates
  - Accurate results
  - Simple layout
- Limitations:
  - Specialized image, music and iTunes duplicate file finders are separate applications.
  - Filters, exclusions and folder duplicate finding are not available in free version.
- Download: <http://www.ashisoft.com/>

# Layout



# Results Page

Result List - Duplicate Files Finder Product Version : 5.4.4.0

Easy Marking
  Delete Duplicates

File Name	Ext	Size	Modified Date	In Folder	Grp
<input checked="" type="checkbox"/> Faron Young - The Complete Capitol Hits of Far...	log	25.02 KB	6/18/2012 12:10 PM	C:\Users\owner\Music\Danny music\What.Cd\The Comple...	1430
<input type="checkbox"/> folder	jpg	298.30 KB	4/22/2011 12:54 PM	C:\Users\owner\Music\Tunes\JAPAN (David Sylvian) - Dis...	1431
<input checked="" type="checkbox"/> japan - adolescent sex a	jpg	298.30 KB	4/22/2011 12:46 PM	C:\Users\owner\Music\Tunes\JAPAN (David Sylvian) - Dis...	1431
<input type="checkbox"/> folder	jpg	377.61 KB	4/22/2011 2:21 PM	C:\Users\owner\Music\Tunes\JAPAN (David Sylvian) - Dis...	1432
<input checked="" type="checkbox"/> japan - quiet life a	jpg	377.61 KB	4/22/2011 1:28 PM	C:\Users\owner\Music\Tunes\JAPAN (David Sylvian) - Dis...	1432
<input type="checkbox"/> folder	jpg	87.74 KB	10/27/2010 3:54 PM	C:\Users\owner\Music\Tunes\jerry lee lewis - all killer, no ...	1433
<input checked="" type="checkbox"/> folder	jpg	87.74 KB	10/27/2010 3:54 PM	C:\Users\owner\Music\Tunes\jerry lee lewis - all killer, no ...	1433
<input type="checkbox"/> Album Art	jpg	14.31 KB	7/17/2011 3:48 AM	C:\Users\owner\Music\Tunes\Me First And The Gimme Gim...	1434
<input checked="" type="checkbox"/> Folder	jpg	14.31 KB	7/17/2011 3:48 AM	C:\Users\owner\Music\Tunes\Me First And The Gimme Gim...	1434
<input type="checkbox"/> [www.PunksAndSkins.com] albumartsmall	jpg	3.55 KB	5/14/2011 3:06 PM	C:\Users\owner\Music\Tunes\NEW YORK DOLLS - [ WWW...	1435
<input checked="" type="checkbox"/> [www.PunksAndSkins.com] albumart_{5a4d08...	jpg	3.55 KB	5/14/2011 3:06 PM	C:\Users\owner\Music\Tunes\NEW YORK DOLLS - [ WWW...	1435
<input checked="" type="checkbox"/> [www.PunksAndSkins.com] albumart_{5a4d08...	jpg	13.42 KB	5/14/2011 3:06 PM	C:\Users\owner\Music\Tunes\NEW YORK DOLLS - [ WWW...	1436
<input type="checkbox"/> [www.PunksAndSkins.com] folder	jpg	13.42 KB	5/14/2011 3:06 PM	C:\Users\owner\Music\Tunes\NEW YORK DOLLS - [ WWW...	1436
<input type="checkbox"/> AlbumArtSmall	jpg	2.41 KB	8/22/2011 7:32 PM	C:\Users\owner\Music\Tunes\Pete Seeger Collection\Pete...	1437
<input checked="" type="checkbox"/> AlbumArt_{4CF3DADE-10EB-42F9-9C8A-2C63D...	jpg	2.41 KB	8/22/2011 7:32 PM	C:\Users\owner\Music\Tunes\Pete Seeger Collection\Pete...	1437
<input checked="" type="checkbox"/> AlbumArt_{4CF3DADE-10EB-42F9-9C8A-2C63D...	jpg	9.64 KB	8/22/2011 7:32 PM	C:\Users\owner\Music\Tunes\Pete Seeger Collection\Pete...	1438
<input type="checkbox"/> Folder	jpg	9.64 KB	8/22/2011 7:32 PM	C:\Users\owner\Music\Tunes\Pete Seeger Collection\Pete...	1438
<input type="checkbox"/> _01 Another Time	mp3	82 Bytes	1/7/2012 1:34 PM	C:\Users\owner\Music\Tunes\Present Tense\_MACOSX\	1439
<input checked="" type="checkbox"/> _02 Song To The Magic Frog (Will You Ever Know)	mp3	82 Bytes	1/7/2012 1:34 PM	C:\Users\owner\Music\Tunes\Present Tense\_MACOSX\	1439



AlbumArtSmall



AlbumArt\_{4CF3DADE-10EB-4...

Total Duplicates: 1690 | Wasted Disk Space: 7.33 GB  
 Files Scan: 38002 | Time Taken: 00:18:31

# MacBooster 2



- Operating System: Mac
- Developer: IObit
- Price: Free demo, \$39.95 for single license
- Features:
  - Tool for security, cleanup, uninstaller, file eraser, and duplicate finder
  - Quick, easy to use
  - Scans multiple locations at once
- Limitations:
  - Free demo is only available for 14 days
- Download: <http://www.macbooster.net/index.php>

# Finds duplicates with different names

The screenshot displays the MacBooster 2 application interface. On the left, a sidebar lists various system optimization tools, with 'Duplicates Finder' highlighted in blue and showing a total of 17.11 GB of duplicates. The main window is divided into a file list on the left and a preview area on the right. The file list is sorted by size and includes a 'Smart check' option. The file 'Rostrevor (40).JPG' is selected and highlighted in blue. The preview area shows a photograph of two women in a field, with the file path and name displayed below it. At the bottom, there are buttons for 'Select All', 'Deselect All', and 'Remove'.

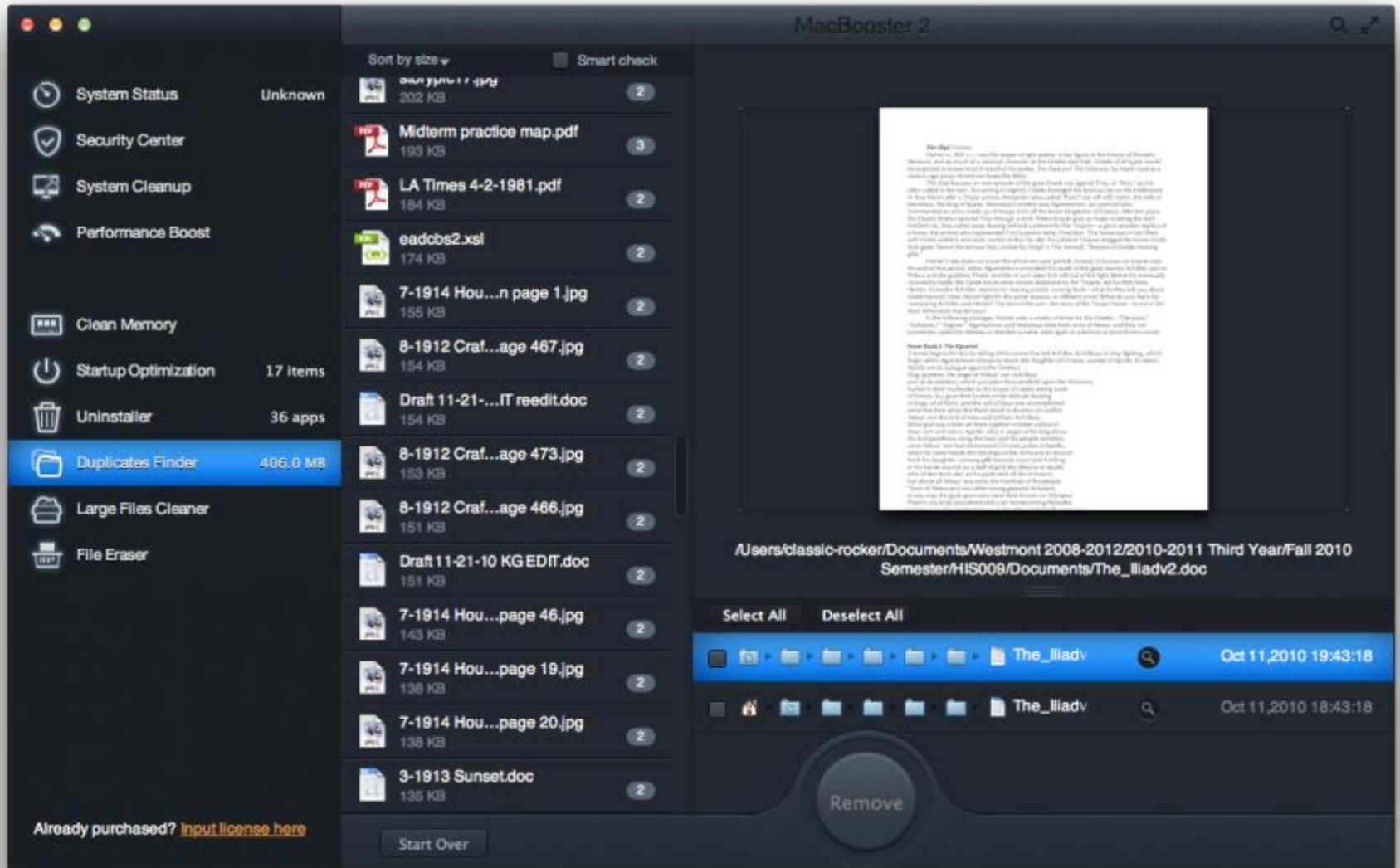
File Name	Size	Count
DSC_0300.JPG	3.6 MB	3
DSC_0880.JPG	3.6 MB	3
DSC_0525.JPG	3.6 MB	2
DSC_0091.JPG	3.6 MB	3
DSC_0420.JPG	3.6 MB	3
<b>Rostrevor (40).JPG</b>	3.6 MB	<b>2</b>
DSC_0508.JPG	3.6 MB	2
DSC_0393.JPG	3.6 MB	2
DSC_0022.JPG	3.6 MB	2
London (22).JPG	3.6 MB	3
DSC_0078.JPG	3.6 MB	2
DSC_0076.JPG	3.6 MB	2
DSC_0664.JPG	3.6 MB	2
DSC_0013.JPG	3.6 MB	2

Path: /Users/classic-rocker/Pictures/09-10 Second Year/\_Europe Semester/9.15-9.20 Rostrevor/DSC\_0617.JPG

Buttons: Select All, Deselect All, Remove

Footer: Already purchased? [Input license here](#)

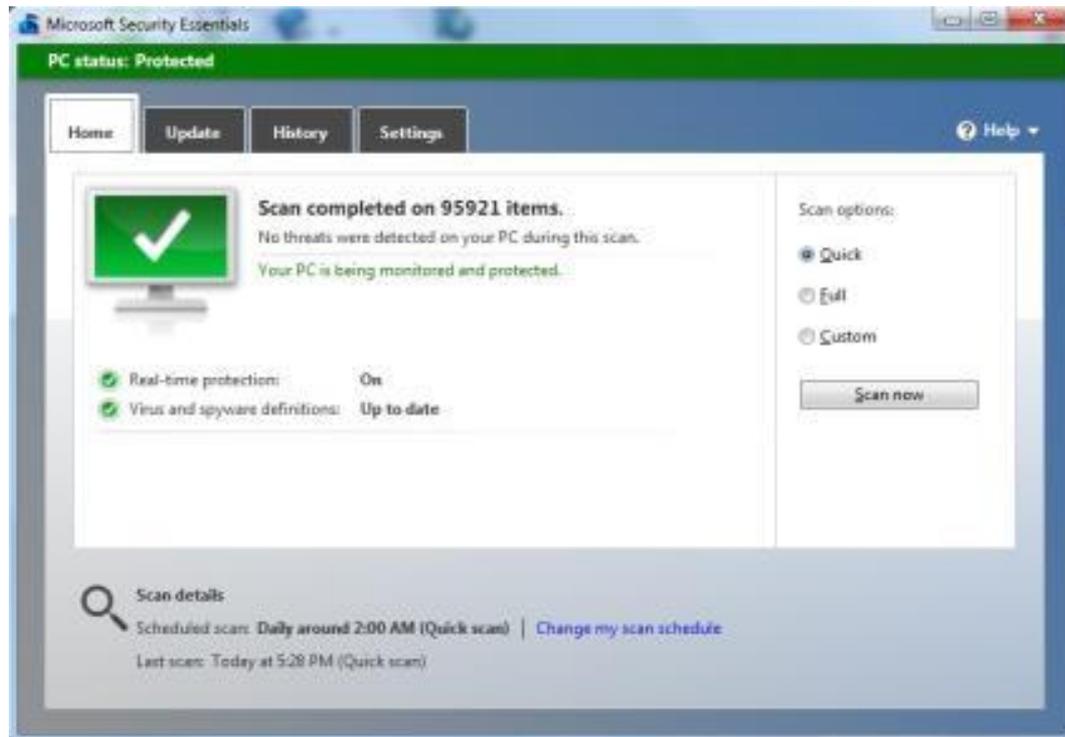
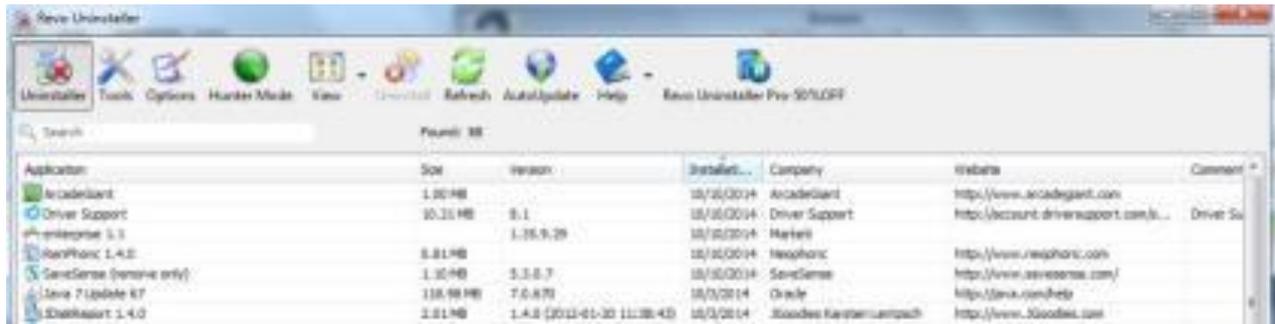
# Multiple file formats



# NoDupe (a cautionary tale)

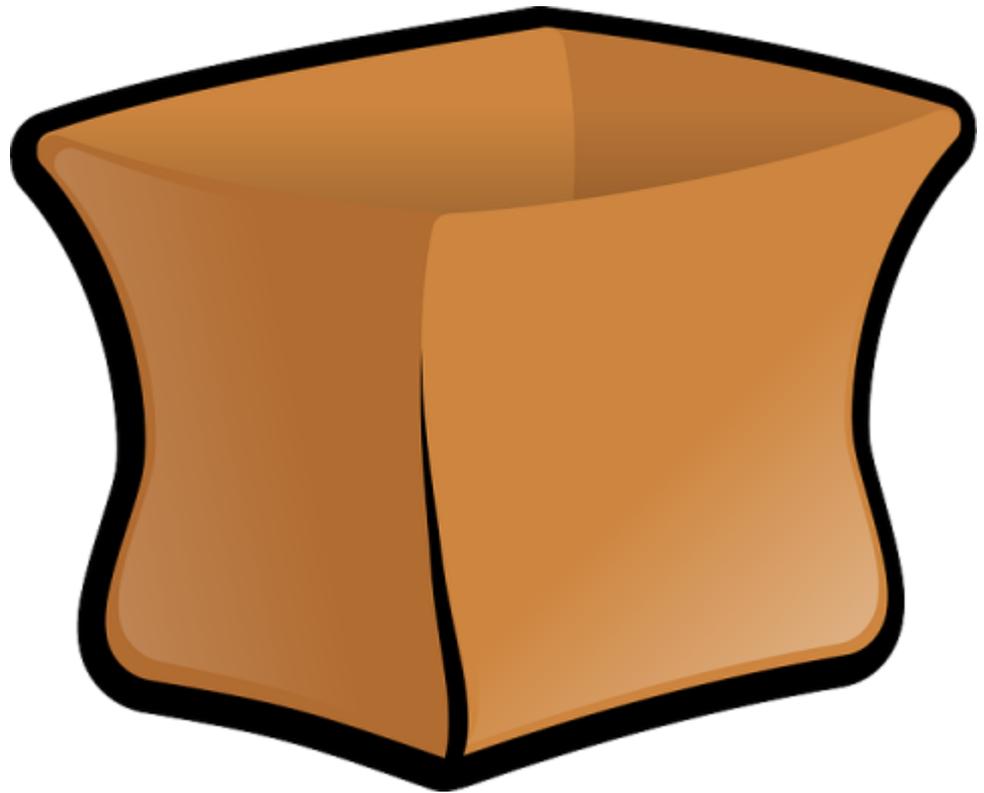
- Operating System: Windows
- Developer: OTISsoft
- Price: Free
- Features:
  - n/a
- Limitations:
  - Adware, bloatware, firefox ad-ons
- Download: Find it yourself

# Revo Uninstaller and MSE Plug



# File transfer

1. BagIt Specification
2. BagIt Library
3. Bagger



# BagIt Specification

- Designed by the Library of Congress
- Developed to support “Transfer” activities:
  - Adding digital content to collections
  - Moving digital content between storage systems
  - Review of digital files for fixity, quality, authoritativeness
  - Inventorying and recording transfer life cycle events for digital files
- Hierarchical file packaging format designed to support the transfer of digital content via media or network
- Specification:  
<http://www.digitalpreservation.gov/documents/bagit-spec.pdf>

# What does BagIt enable?

- Allows digital content to be packaged in “bags” with metadata and a manifest that details checksums so that contents can be verified upon receipt of bags
- A “bag” includes:
  - A data directory
  - A manifest file
  - An information file

```
<bag_dir>/  
| manifest-<algorithm>.txt  
| bagit.txt  
| [optional additional tag files]  
\--- data/  
    | [optional file hierarchy]
```

# Example Uses of BagIt

- Archivematica – create OAIS Archival Information Packages (AIP)
- Ghent University – archival format for digital collections and interchange format
- Dryad Data Repository – share data and related metadata with TreeBASE
- Stanford Digital Repository – primary transfer format
- University of North Texas – archival container format in digital repository, interchange format

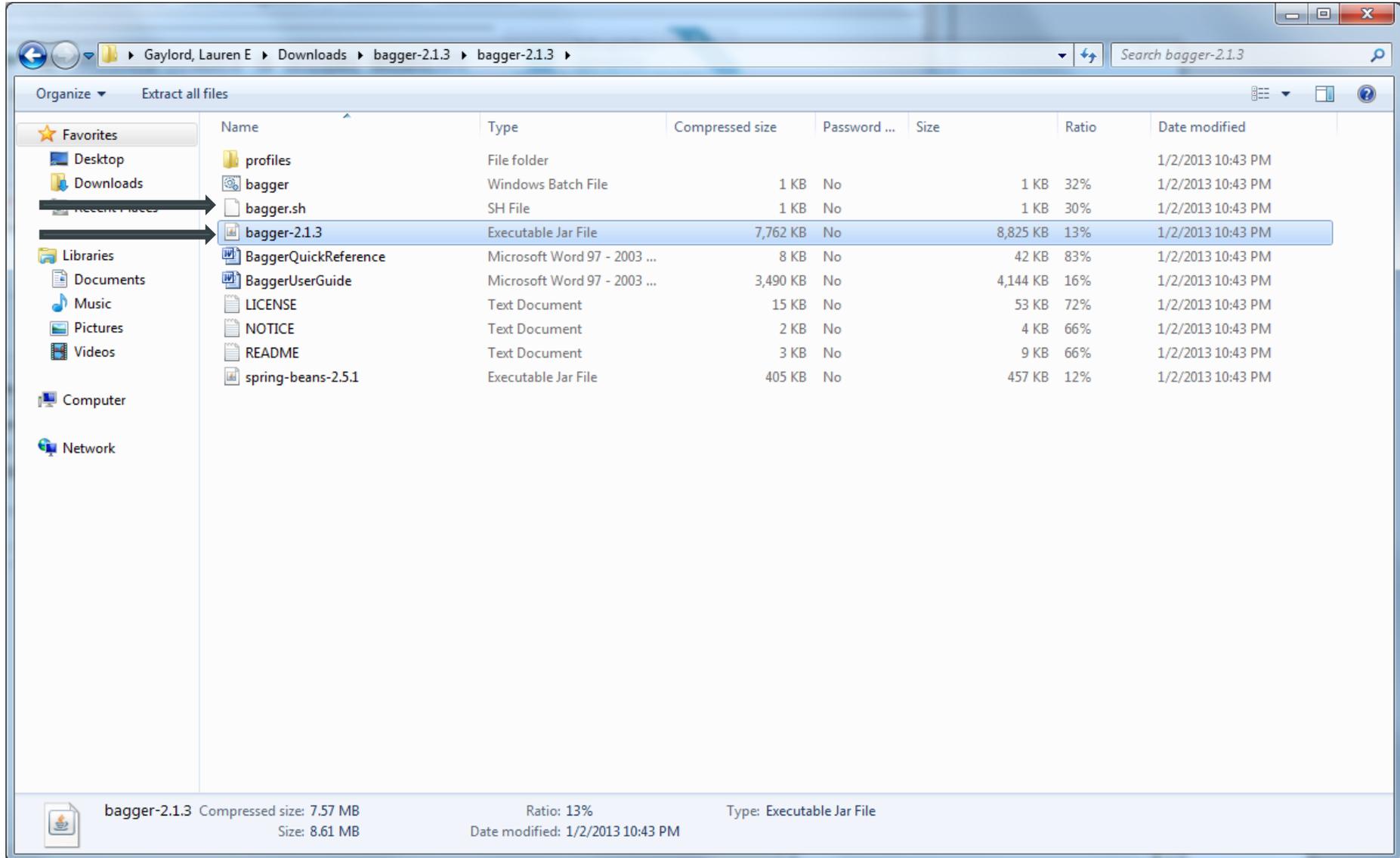
# BagIt tools – BagIt Library (BIL)

- Operating System: Windows, Linux
- Developer: Library of Congress
- Description: Software library intended to support the creation, manipulation, and validation of bags.  
Command-line driver
- Features:
  - Creating, manipulating, validating, verifying Bags, reads from and writes to number of formats
- Limitations:
  - Sparse documentation
  - Command-line
- Download:  
<https://github.com/LibraryOfCongress/bagit-java>

# BagIt Tools - Bagger

- Operating System: Windows, Linux
- Developer: Library of Congress
- Description: Graphical User Interface for BIL
- Features:
  - Provides visual view of bag contents, bag state, options
  - More extensive documentation
- Limitations:
  - Possibly not updated as LoC has migrated to github to distribute software
- Download: <http://sourceforge.net/projects/loc-xferutils/files/loc-bagger/2.1.3/>

# Installing Bagger

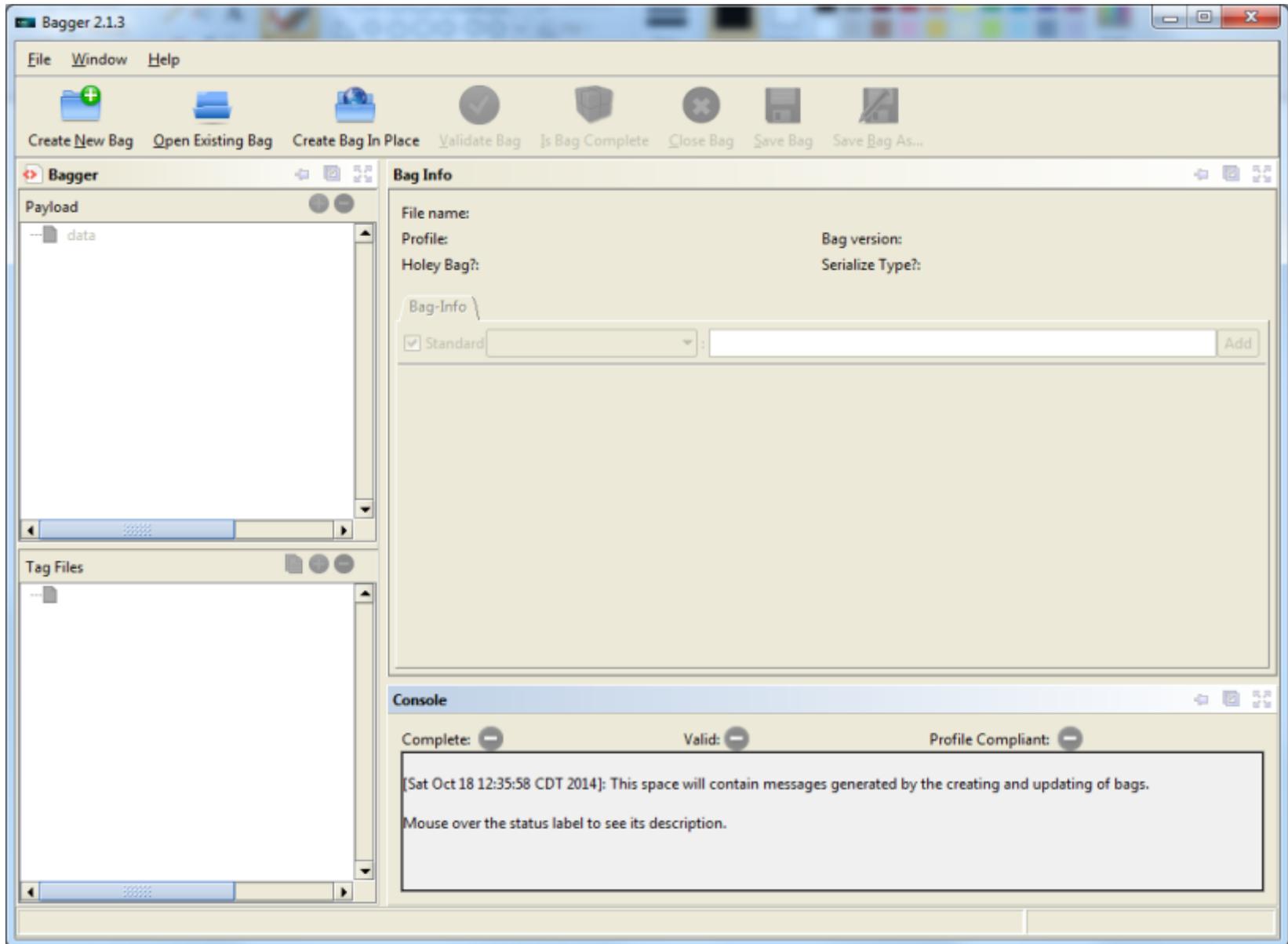


The screenshot shows a Windows Explorer window with the address bar displaying the path: `Gaylord, Lauren E > Downloads > bagger-2.1.3 > bagger-2.1.3`. The window title is `Search bagger-2.1.3`. The main pane shows a list of files and folders extracted from a zip archive. The file `bagger-2.1.3` is selected, and its details are shown in the status bar at the bottom.

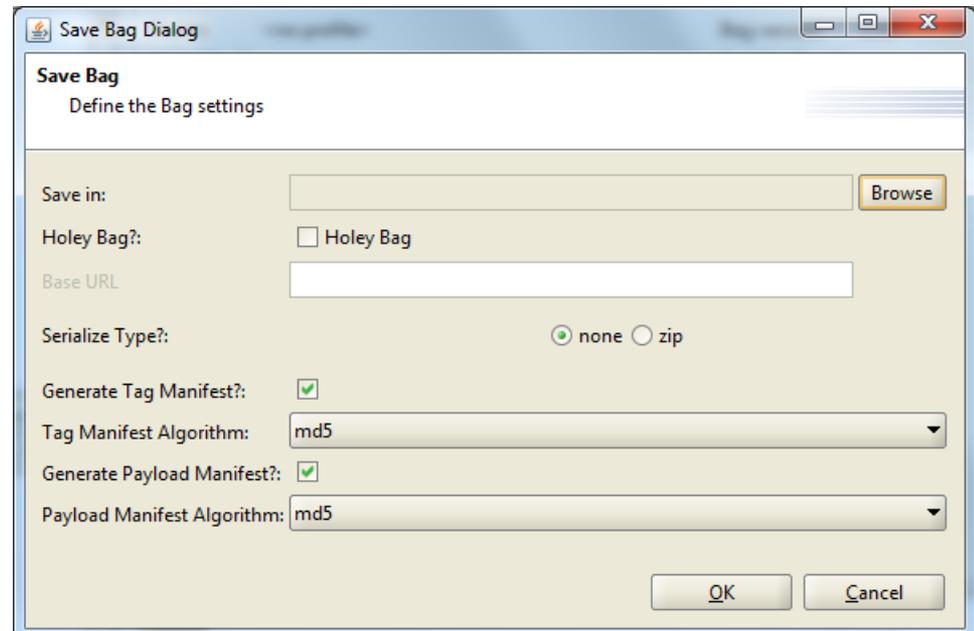
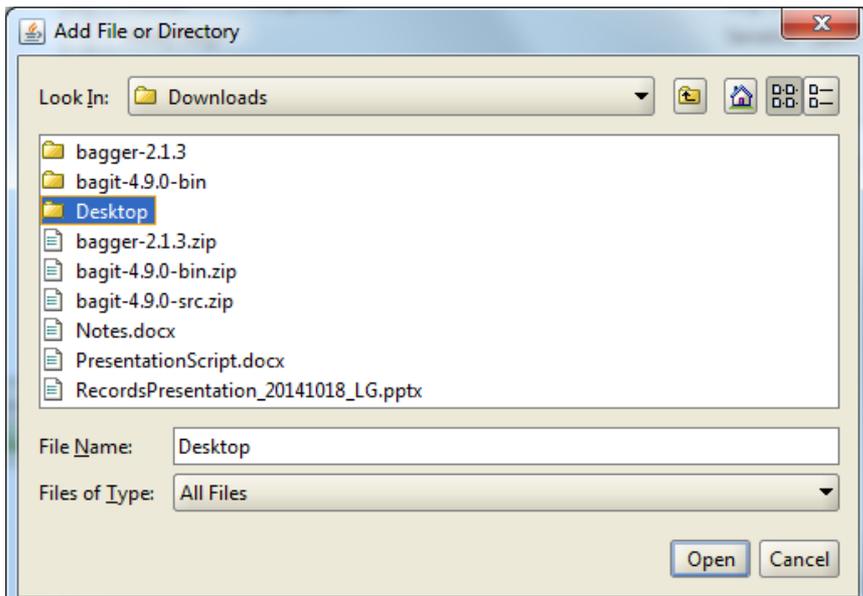
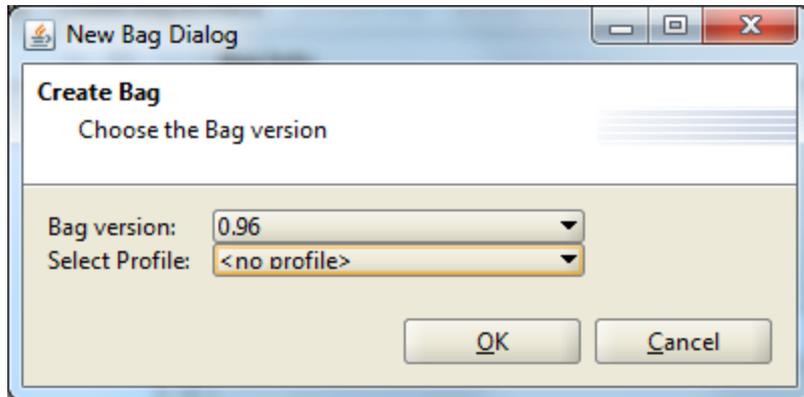
Name	Type	Compressed size	Password ...	Size	Ratio	Date modified
profiles	File folder					1/2/2013 10:43 PM
bagger	Windows Batch File	1 KB	No	1 KB	32%	1/2/2013 10:43 PM
bagger.sh	SH File	1 KB	No	1 KB	30%	1/2/2013 10:43 PM
<b>bagger-2.1.3</b>	<b>Executable Jar File</b>	<b>7,762 KB</b>	<b>No</b>	<b>8,825 KB</b>	<b>13%</b>	<b>1/2/2013 10:43 PM</b>
BaggerQuickReference	Microsoft Word 97 - 2003 ...	8 KB	No	42 KB	83%	1/2/2013 10:43 PM
BaggerUserGuide	Microsoft Word 97 - 2003 ...	3,490 KB	No	4,144 KB	16%	1/2/2013 10:43 PM
LICENSE	Text Document	15 KB	No	53 KB	72%	1/2/2013 10:43 PM
NOTICE	Text Document	2 KB	No	4 KB	66%	1/2/2013 10:43 PM
README	Text Document	3 KB	No	9 KB	66%	1/2/2013 10:43 PM
spring-beans-2.5.1	Executable Jar File	405 KB	No	457 KB	12%	1/2/2013 10:43 PM

**bagger-2.1.3** Compressed size: 7.57 MB  
Size: 8.61 MB  
Ratio: 13%  
Date modified: 1/2/2013 10:43 PM  
Type: Executable Jar File

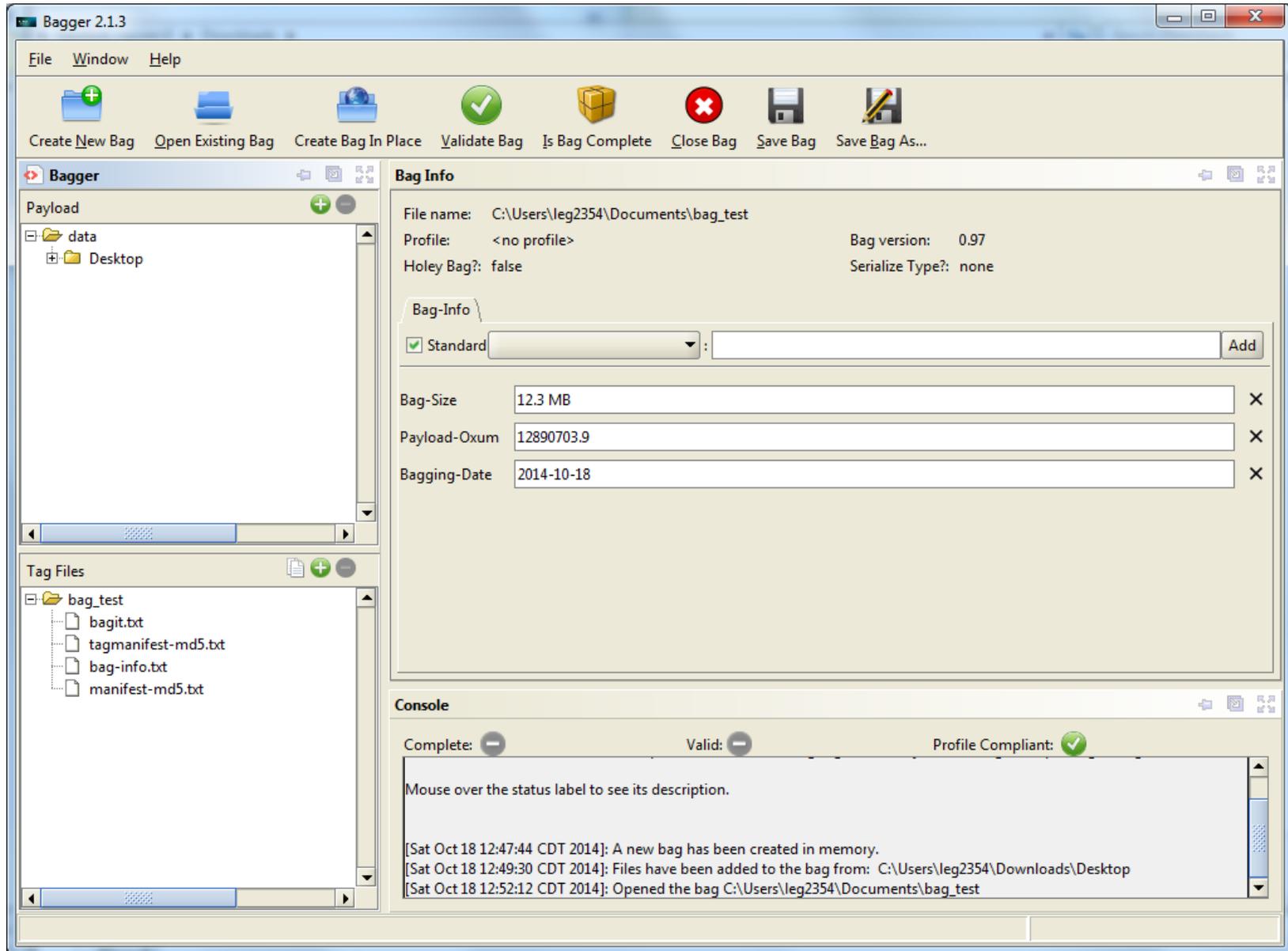
# Bagger Main Window



# Creating and Saving a Bag



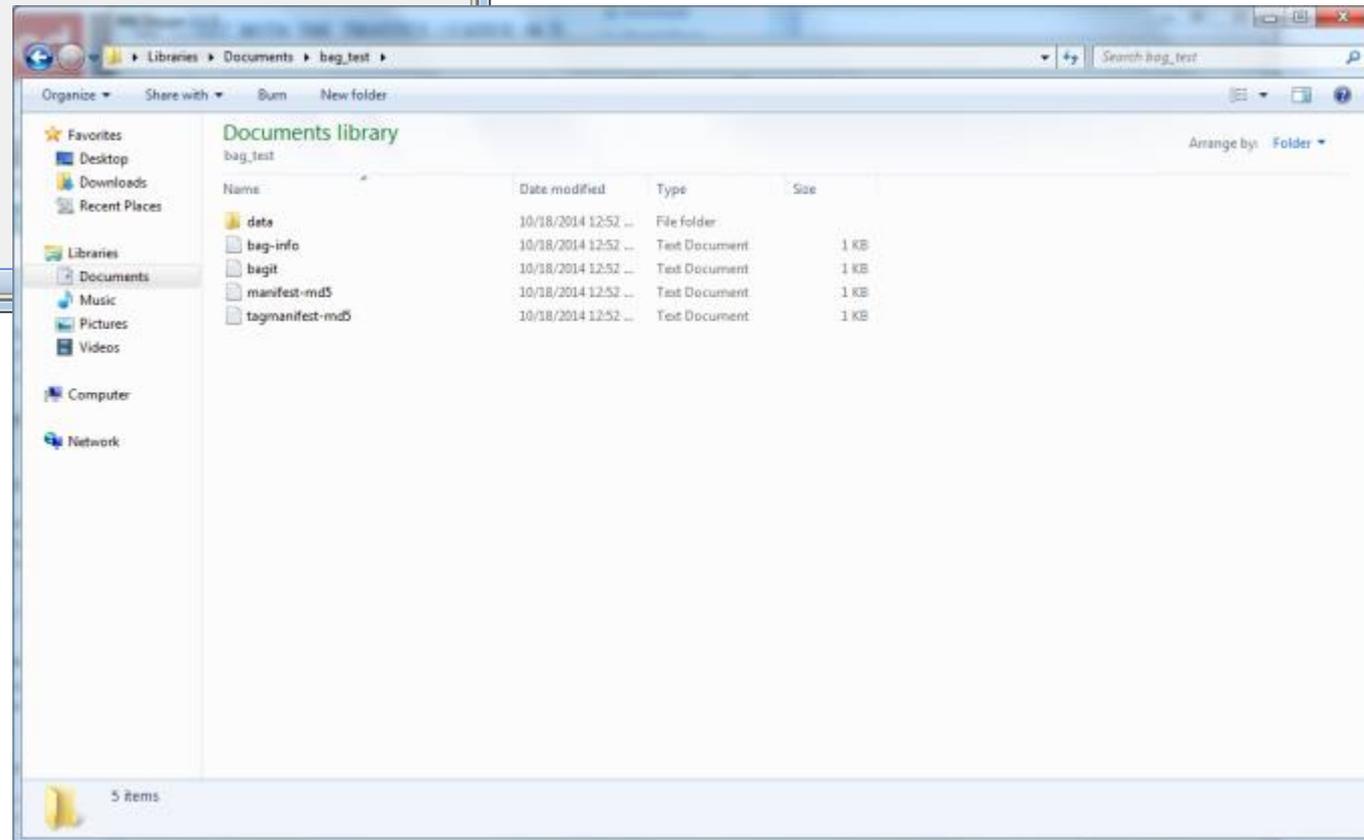
# Saved Bag



# Checksums and saved bag on disk

```
View Tags
bagit.txt tagmanifest-md5.txt bag-info.txt manifest-md5.txt data/

{data/Desktop/Pictures/Desert.jpg=ba45c8f60456a672e003a875e469d0eb,
data/Desktop/Pictures/Animals/Koala.jpg=2b04df3ecc1d94afddff082d139c6f15,
data/Desktop/Pictures/Lighthouse.jpg=8969288f4245120e7c3870287cce0ff3,
data/Desktop/Pictures/Animals/Penguins.jpg=9d377b10ce778c4938b3c7e2c63a229a,
data/Desktop/Pictures/Hydrangeas.jpg=bdf3bf1da3405725be763540d6601144,
data/Desktop/Pictures/Tulips.jpg=fafa5eafeaf3cbe3b23b2748d13e629a1,
data/Desktop/Documents/RecordsPresentation_20141018_LG.pptx=5264eaf5f6a80f838d3ea,
data/Desktop/Pictures/Chrysanthemum.jpg=076e3caed758a1c18c91a0e9cae3368f,
data/Desktop/Pictures/Animals/Jellyfish.jpg=5a44c7ba5bbe4ec867233d67e4806848}
```



# Tools for various programming environments

- [Archive::BagIt](#): Perl
- [BagIt Library](#): Java
- [BagIt gem](#): Ruby
- [bagit](#): Python
- [pybagit](#): Python
- [BagIt GUI](#): JRuby
- [BagItPHP](#): PHP

# Conclusions & Implications

- File re-naming and de-duplicating software are, for the most part, easy to use. Useful for any computer user who stores files on his/her computer.
- File transfer programs have a higher learning curve and require a good reason (i.e. archival) and decent understanding of technical environment before using.
- De-Duplicating can have very serious consequences for your files. Make sure you know what you're doing.

# Resources

## File Renaming and De-Duplicating:

- <http://download.cnet.com/>
- [www.wikipedia.com](http://www.wikipedia.com) (for holistic overview)
- [www.pcworld.com/](http://www.pcworld.com/)
- [www.macworld.com/](http://www.macworld.com/)
- <http://www.iso.org/>

## BagIt/Bagger

- <http://www.digitalpreservation.gov/multimedia/videos/bagit0609.html>
- [http://www.ncdcr.gov/Portals/26/PDF/guidelines/Using\\_Bagger.pdf](http://www.ncdcr.gov/Portals/26/PDF/guidelines/Using_Bagger.pdf)
- [http://project.wdl.org/arab\\_peninsula/workshop2012/en/doha\\_workshop\\_2012\\_bagger\\_en.pdf](http://project.wdl.org/arab_peninsula/workshop2012/en/doha_workshop_2012_bagger_en.pdf)

Questions?