DupLichaSe v0.1.**

(Simple Duplicate File detector with Semi-automatic Suggestion System)

The Dirty MANUAL . . .v2 ---> by Timothy

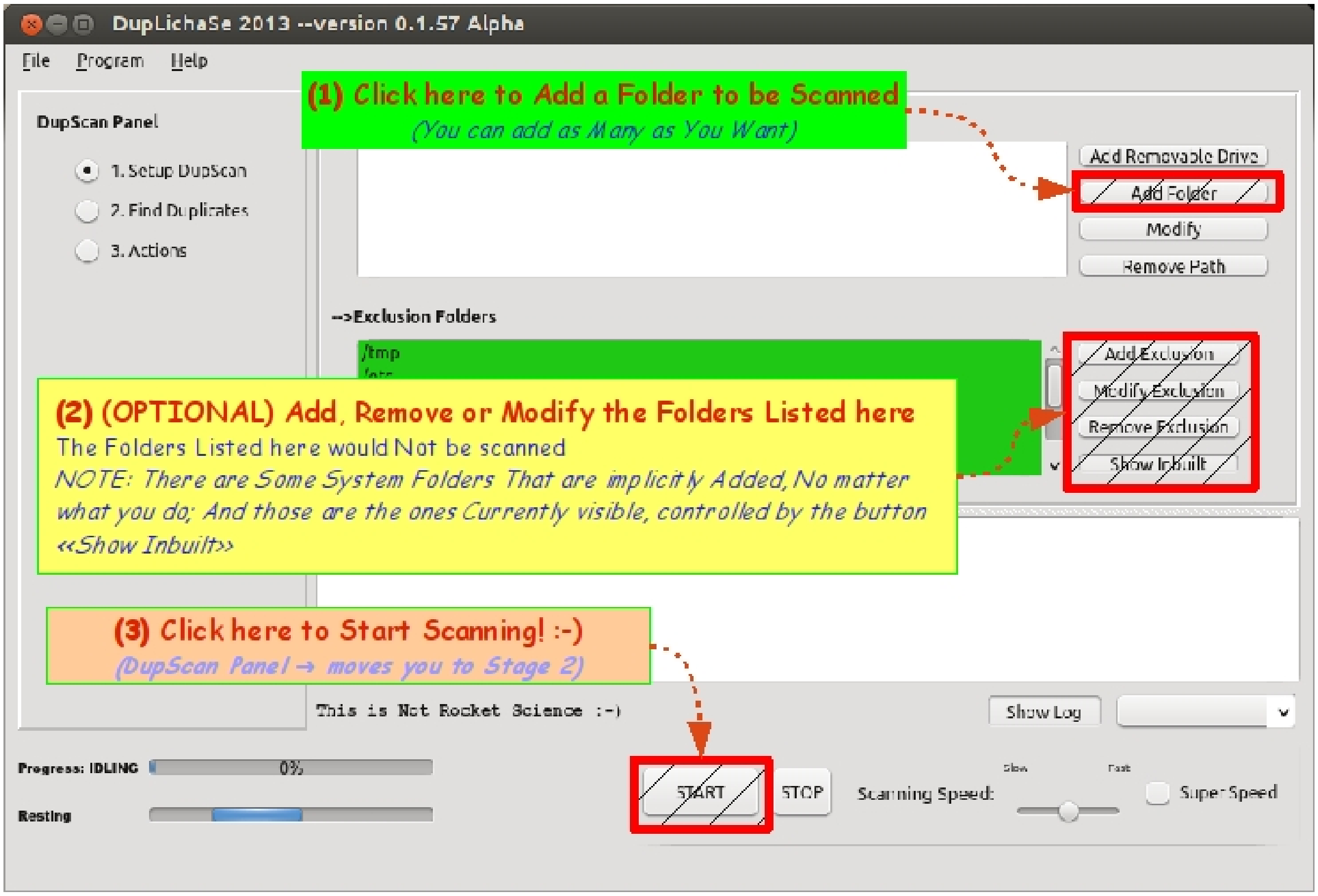
in Memory of my Late Dad, "Prof. M. I. Onogu" who passed away on 6th May, 2013

<<Note This is NOT a comprehensive MANUAL>> It was written under 30 mins!! just to meet a deadline... please refer to http://www.sourceforge.net/projects/duplichase *for up to date Manual or my blog...*

1 Minute Tutorial

Stage 1 → Setup

The Software starts with this Window, this is where you must select Folders to be scanned for Duplicate files.



(1) Click here to Add a Folder to be Scanned

(You can add as Many as You Want)

(2) (OPTIONAL) Add, Remove or Modify the Folders Listed here

The Folders Listed here would Not be scanned

NOTE: There are Some System Folders That are implicitly Added, No matter what you do; And those are the ones Currently visible, controlled by the button <<Show Inbuilt>>

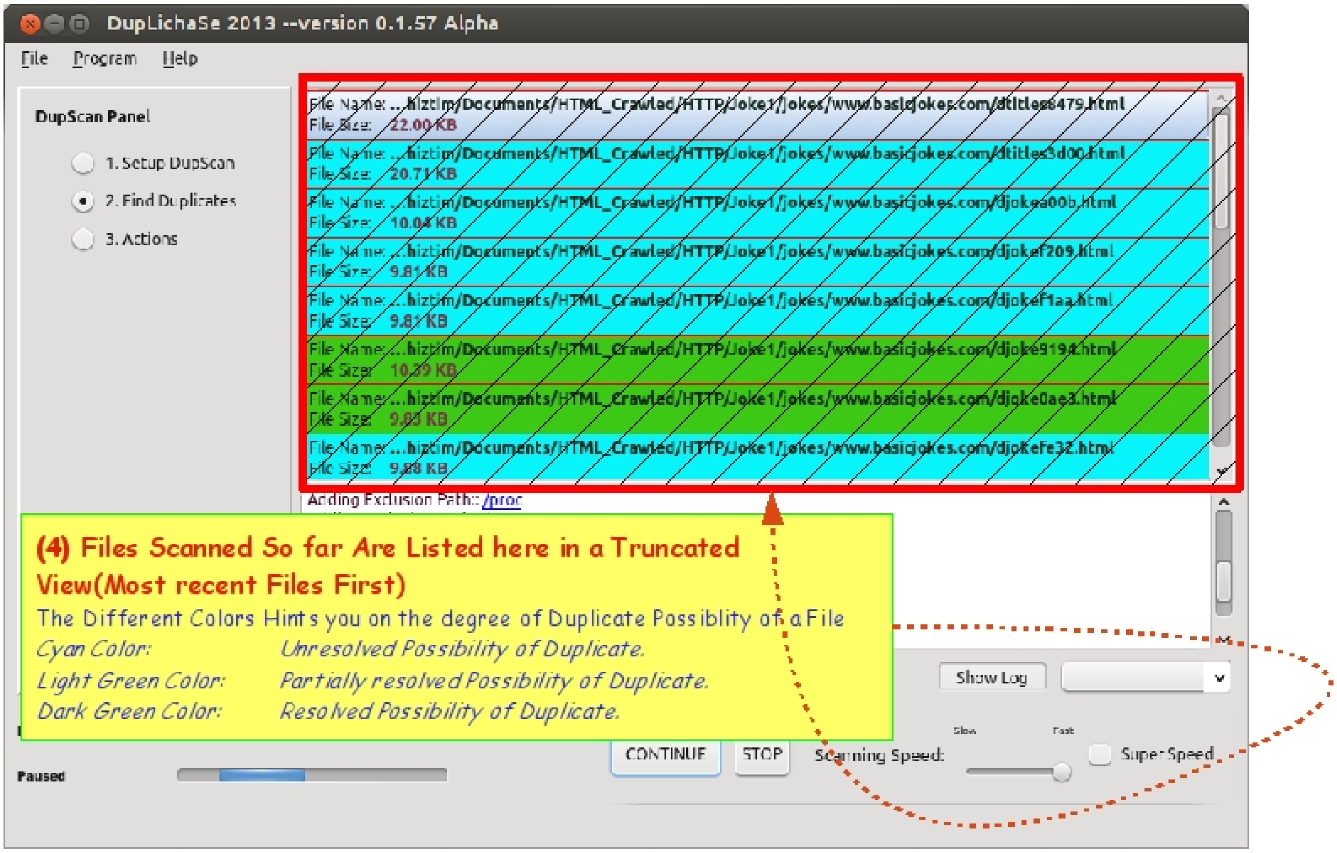
(3) Click here to Start Scanning! :-)

(DupScan Panel → moves you to Stage 2)

This is Not Rocket Science :-)

Stage 2 → Find Duplicates

After clicking start, DupScan automatically changes the view to **(2) Find Duplicate** as shown below... Progress can be viewed in RealTime*(slower)*;... or off realtime*(faster)* by clicking on **Super Speed**. Speed can also be variated to lessen or increase the scan speed at the expense of CPU power and Disk access



```
File Name: ...hiztim/Documents/HTML_Crawled/HTTP/Joke1/jokes/www.basicjokes.com/dtitles8479.html
File Size:   22.00 KB
File Name: ...hiztim/Documents/HTML_Crawled/HTTP/Joke1/jokes/www.basicjokes.com/dtitles3d00.html
File Size:   20.71 KB
File Name: ...hiztim/Documents/HTML_Crawled/HTTP/Joke1/jokes/www.basicjokes.com/djokea00b.html
File Size:   10.04 KB
File Name: ...hiztim/Documents/HTML_Crawled/HTTP/Joke1/jokes/www.basicjokes.com/djokef209.html
File Size:   9.81 KB
File Name: ...hiztim/Documents/HTML_Crawled/HTTP/Joke1/jokes/www.basicjokes.com/djokef1aa.html
File Size:   9.81 KB
File Name: ...hiztim/Documents/HTML_Crawled/HTTP/Joke1/jokes/www.basicjokes.com/djoke9194.html
File Size:   10.39 KB
File Name: ...hiztim/Documents/HTML_Crawled/HTTP/Joke1/jokes/www.basicjokes.com/djoke0ae3.html
File Size:   9.83 KB
File Name: ...hiztim/Documents/HTML_Crawled/HTTP/Joke1/jokes/www.basicjokes.com/djokefe32.html
File Size:   9.88 KB
Adding Exclusion Path:: /proc
```

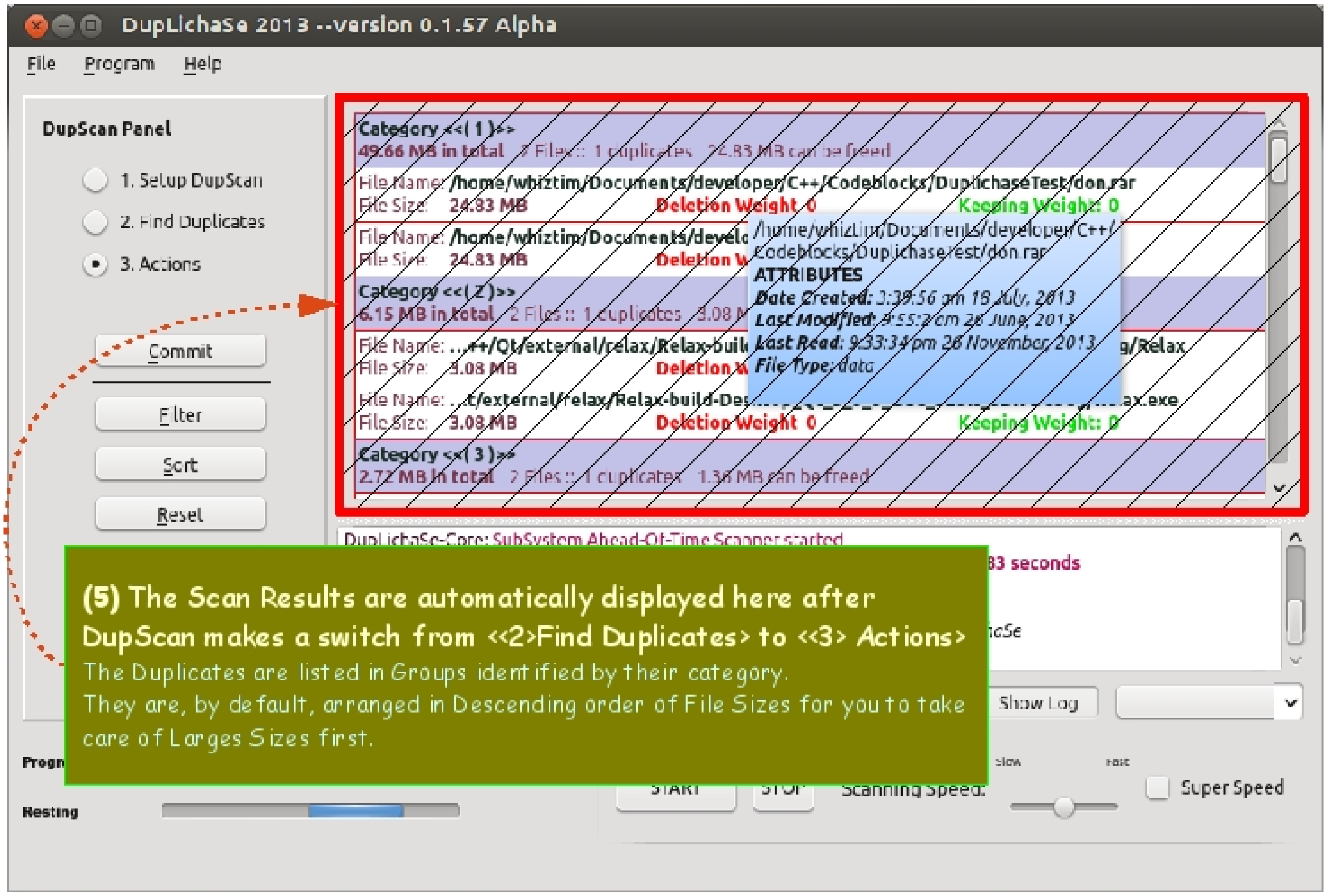
(4) Files Scanned So far Are Listed here in a Truncated View(Most recent Files First)

The Different Colors Hints you on the degree of Duplicate Possiblity of a File

Cyan Color:	*Unresolved Possibility of Duplicate.*
Light Green Color:	*Partialy resolved Possibility of Duplicate.*
Dark Green Color:	*Resolved Possibility of Duplicate.*

Stage 2 → Actions

After scanning is complete, DupScan automatically changes the view to **(3) Actions** as shown below... This is a very intuitive and non interfering interface; Here, you can take all your decisions concerning your file; from, deleting to Recycle Bin or Automatic Selections...



```
Category <<( 1 )>>
49.66 MB in total   2 Files :: 1 duplicates   24.83 MB can be freed
File Name: /home/whiztim/Documents/developer/C++/Codeblocks/DuplichaseTest/don.rar
File Size:   24.83 MB              Deletion Weight  0              Keeping Weight:  0
File Name: /home/whiztim/Documents/develo
File Size:   24.83 MB              Deletion W
Category <<( 2 )>>
6.15 MB in total   2 Files :: 1 duplicates   3.08 M
File Name: ...++/Qt/external/relax/Relax-buil                                         g/Relax
File Size:   3.08 MB               Deletion W
File Name: ...t/external/relax/Relax-build-Des                                         ax.exe
File Size:   3.08 MB               Deletion Weight  0              Keeping Weight:  0
Category <<( 3 )>>
2.72 MB in total   2 Files :: 1 duplicates   1.36 MB can be freed
```

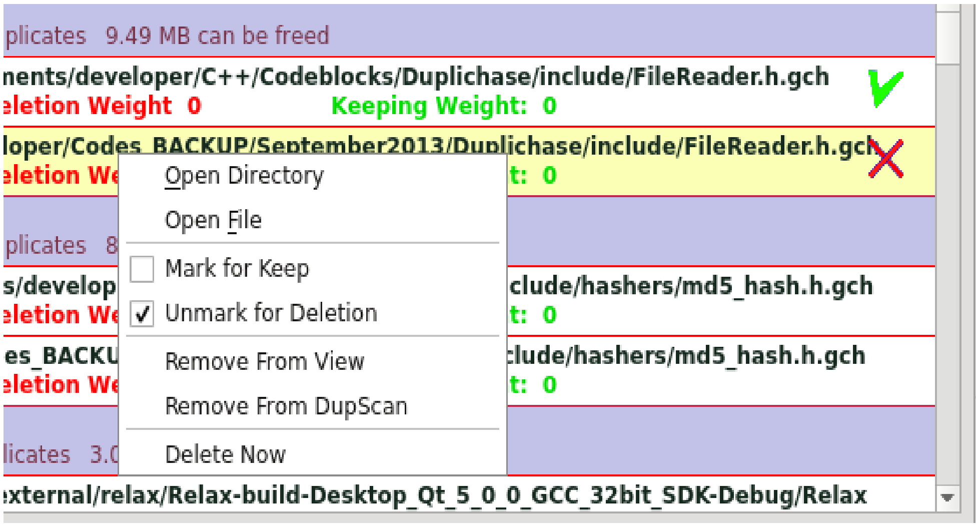
```
/home/whiztim/Documents/developer/C++/
Codeblocks/DuplichaseTest/don.rar
ATTRIBUTES
Date Created: 3:39:56 pm 18 July, 2013
Last Modified: 9:55:2 am 26 June, 2013
Last Read: 9:33:34 pm 26 November, 2013
File Type: data
```

```
DupLichaSe-Core: SubSystem Ahead-Of-Time Scanner started
33 seconds
iaSe
```

(5) The Scan Results are automatically displayed here after DupScan makes a switch from <<2>Find Duplicates> to <<3> Actions>

The Duplicates are listed in Groups identified by their category.

They are, by default, arranged in Descending order of File Sizes for you to take care of Larges Sizes first.



```
plicates   9.49 MB can be freed
nents/developer/C++/Codeblocks/Duplichase/include/FileReader.h.gch   ✓
eletion Weight  0          Keeping Weight:  0
loper/Codes_BACKUP/September2013/Duplichase/include/FileReader.h.gch   ✗
eletion We                                      t:  0
plicates   8
s/develop                                       clude/hashers/md5_hash.h.gch
eletion We                                      t:  0
es_BACKU                                        clude/hashers/md5_hash.h.gch
eletion We                                      t:  0
licates   3.0
xternal/relax/Relax-build-Desktop_Qt_5_0_0_GCC_32bit_SDK-Debug/Relax
```

BASIC ACTIONS!

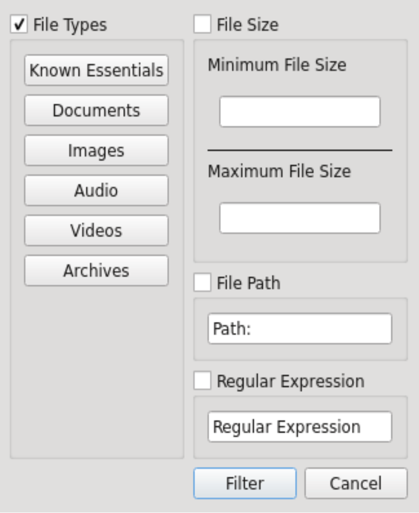
Now, right click on any file and the options shown below are listed for you... Choose **Delete Now** ..to get rid of it!!...

CONGRATULATIONS!

The ***recommender System*** will learn from your actions and update the *Deletion Weight* and *Keeping Weight*.

Actually, these two weights are supposed to be normalized and hidden from you! . . . but
+<1> for current resources including time... There is no Normalization.. so the values will forever go up until about “1800000000000000000”...before DupLichaSe may crash! :(.
+<2> for geeky sake, the values are shown to you!

Filter:



Brief: These options are presented for filtering and should be used with care... to show only files that meet to the selected filtering criteria...

NOTE: when you notice any funny stuff in the view, click **Reset** and possibly save your file {and if you care enough, report the issue}

++>**File Types:**
click one or more file type categorization;
Known Essentials, currently is implemented as an aggregate of popular file types;

++>**File size:**
should input the sizes with prefixes example: **45KB, 1.322GB, 23453B** are all acceptable and conversions will be appropriately made by the underlying subsystem.

++>**File Path:**
should show only the files that are in the path inputed.

++>**Regular Expression:**
Show only files that matches the PERL standard Regex pattern.

F A Q

DupLichaSe just crashed what happend?

Oopsie!! I deeply apologize for that. Though That's naturally what you should expect from Software in Alpha releases... The DupLichaSe is still undergoing some refinements and bug fixes and testing. DupLichaSe has somewhat close to 11, 000 lines of code... about 9,000+ useful.. And it's not so easy. *It's worth mentioning that I initially started writing this Software* ***as a toy program*** *to experiment with a few C++11 features and at the same time remove duplicate files from my System... ..but due to a few events in my Life, I decided to take it make it usable for other non-techies... So, by my standards, it's Code base is* ***horrible*** well, I have been doing some refactoring work... **Mind you, I only work on DupLichaSe in my spare time, which is hard to come by. BUT it shouldn't prevent you from reporting issues**

Why does DupLichaSe tend to hang sometimes?

Most times, the slow rate of reading files from Hard disk is the cause... Larger files will halt DupLichaSe's scan progress for longer periods than smaller files. On the other hand **There will always be a NON-RESPONSIVE delay after scanning has finished, just before Action view is shown... The delay time is proportional to the amount of duplicate Files**. This may be between <1sec to minutes

Can I get False positives from DupLichaSe?

Yes, but That is a very very low probability. . . which is even dependent on the File Size, here is a table that summarizes a quick and dirty probability distribution(not statistically accurate, but a blind estimate).. In reality, you only need to watch out for small sized files, (files less than 10KB)....

S/N	File Size	False positive probability (average)
1	1byte ~ 1KB	0.00000976562
2	1KB ~ 10KB	0.0000000244141
3	10KB ~ 100KB	0.000000000007
4	100KB ~ 2MB	0.000000000000000847
5	2M ~ 100MB	0.000000000132
6	100MB+	0.000000000323

Can I undelete any file deleted with DupLichaSe?

Yes, you can. AS a matter of principle, DupLichaSe is VERY CAREFUL with EVERYFILE it encounters. **Any File deleted with via DupLichaSe goes to the Recycle bin or trash...** if it cannot, the file will NOT be deleted!.. It is your responsibility to delete them from your trash can!

What detection algorithm does DupLichaSe use?

A very simple inductive heuristics... Let **A** and **B** represent two files being compared by the Software.... here is what happens

- if the *file size of* **A** is equal to the *file size of* **B**, then **A** and **B** are possible duplicates.
- Then if the *hash value of the first (x)% of a forward read of* **A** is equal to the *hash value of the first (x)% of a forward read of* **B**, then **A** and **B** have increased chances of being duplicates. Where (x) varies from 100 to 10 and is dependent on file sizes...
- Then if the *hash value of the first (y)% of a backward read* **A** is equal to the *hash value of the first (y)% of a backward read of* **B**, then **A** and **B** have extremely high chances being duplicates. (more than (99%). Where (y) varies from 100 to 20 and is dependent on file sizes...
- Lastly, binary comparison, which is currently disabled... but a tool may be provided for that

ACKNOWLEDGEMENT

There is a non exhaustive list of people that must be mentioned who were directly or indirectly instrumental to my life endeavors, passion and sanity. The is only, but, a few of the people I did love to acknowledge.
My profound appreciation goes to my ever strong and loving Mom, Veronica Onogu, whom words are too brief to describe.

My late Dad, M. I. Onogu who was a Professor and a former veteran of Electrical, Electronics and Control Systems Engineering who taught me some undergraduate Mathematics at a very young age and whom unfortunately, passed away on 6th May, 2013.

My Sisters who have been very supportive.

My friends: but not limited to;
+Ahmed Muhammad Bello;
+Daser Solomon Retnan;
+Olabode Peter Hotonu;
+Haruna Manzo Manzo;
+Emmanuel Idusuyi;
+Joshua Mabadeje;
+Aderonke Omole;
+Onmikpa Ogah;
+Halima Umar;
+Ebi Laffin;

...and a host of others who will have to forgive me for omitting their names.

My Instructors and Lecturers at the Mechanical and Production Engineering A. T. B. U Bauchi, and neighbor, the entirety of Prof. E. B. Agbo's family.

Finally but not religiously,
I acknowledge and appreciate God whom has given me this gift of Life, purpose, and passion.
I acknowledge and appreciate Jesus Christ of whom has given me the best gift in Life, FREE access to the throne of God.
I acknowledge and appreciate my direct Boss, the Holy Spirit whom has been my guide, friend.