

# Dublettensuche unter GNU/Linux

Edgar 'Fast Edi' Hoffmann

Community FreieSoftwareOG

*[kontakt@freiesoftwareog.org](mailto:kontakt@freiesoftwareog.org)*

6. November 2019

# Dubletten Begriffserklärung

# Dubletten

## Begriffserklärung

Für den Begriff "Dubletten" gibt es lt. Wikipedia mehrere Definitionen.  
Eine davon bezieht sich auf die IT:

# Dubletten

## Begriffserklärung

Für den Begriff "Dubletten" gibt es lt. Wikipedia mehrere Definitionen.

Eine davon bezieht sich auf die IT:

- Ein mehrfach vorhandener Datensatz einer Datenbank, der redundant, d. h. mehrfach, vorhanden ist

# Dubletten

## Begriffserklärung

Für den Begriff "Dubletten" gibt es lt. Wikipedia mehrere Definitionen.

Eine davon bezieht sich auf die IT:

- Ein mehrfach vorhandener Datensatz einer Datenbank, der redundant, d. h. mehrfach, vorhanden ist

# Dubletten

## Begriffserklärung

Für den Begriff "Dubletten" gibt es lt. Wikipedia mehrere Definitionen.

Eine davon bezieht sich auf die IT:

- Ein mehrfach vorhandener Datensatz einer Datenbank, der redundant, d. h. mehrfach, vorhanden ist

Im heutigen Thema geht es aber um doppelt vorhandene Dateien.

# Dublettensuche Software

# Dublettensuche Software

- FSlint

# Dublettensuche Software

- FSlint
- DupeGuru

# Dublettensuche Software

- FSlint
- DupeGuru
- fdupes

# Dublettensuche FSlint

# Dublettensuche

## FSlint

FSlint ist eine Sammlung nützlicher Hilfsprogramme zum Aufräumen der Festplatte, die unter einer gemeinsamen grafischen Oberfläche erreichbar sind.

Dabei steht FS für FileSystem und lint für Staub.

# Dublettensuche

## FSlint

FSlint ist eine Sammlung nützlicher Hilfsprogramme zum Aufräumen der Festplatte, die unter einer gemeinsamen grafischen Oberfläche erreichbar sind.

Dabei steht FS für FileSystem und lint für Staub.

Mit ihnen lassen sich u.a. folgende Dinge auf der Festplatte aufspüren:

- doppelte Dateien (nützlich z.B. bei großen Musiksammlungen)

# Dublettensuche

## FSlint

FSlint ist eine Sammlung nützlicher Hilfsprogramme zum Aufräumen der Festplatte, die unter einer gemeinsamen grafischen Oberfläche erreichbar sind.

Dabei steht FS für FileSystem und lint für Staub.

Mit ihnen lassen sich u.a. folgende Dinge auf der Festplatte aufspüren:

- doppelte Dateien (nützlich z.B. bei großen Musiksammlungen)
- leere Verzeichnisse

# Dublettensuche

## FSlint

FSlint ist eine Sammlung nützlicher Hilfsprogramme zum Aufräumen der Festplatte, die unter einer gemeinsamen grafischen Oberfläche erreichbar sind.

Dabei steht FS für FileSystem und lint für Staub.

Mit ihnen lassen sich u.a. folgende Dinge auf der Festplatte aufspüren:

- doppelte Dateien (nützlich z.B. bei großen Musiksammlungen)
- leere Verzeichnisse
- Dateien mit nicht UTF8-konformen Namen

# Dublettensuche

## FSlint

FSlint ist eine Sammlung nützlicher Hilfsprogramme zum Aufräumen der Festplatte, die unter einer gemeinsamen grafischen Oberfläche erreichbar sind.

Dabei steht FS für FileSystem und lint für Staub.

Mit ihnen lassen sich u.a. folgende Dinge auf der Festplatte aufspüren:

- doppelte Dateien (nützlich z.B. bei großen Musiksammlungen)
- leere Verzeichnisse
- Dateien mit nicht UTF8-konformen Namen
- temporäre Dateien

# Dublettensuche

## FSlint

FSlint ist eine Sammlung nützlicher Hilfsprogramme zum Aufräumen der Festplatte, die unter einer gemeinsamen grafischen Oberfläche erreichbar sind.

Dabei steht FS für FileSystem und lint für Staub.

Mit ihnen lassen sich u.a. folgende Dinge auf der Festplatte aufspüren:

- doppelte Dateien (nützlich z.B. bei großen Musiksammlungen)
- leere Verzeichnisse
- Dateien mit nicht UTF8-konformen Namen
- temporäre Dateien
- fehlerhafte symbolische Verknüpfungen

# Dublettensuche

## FSlint

FSlint ist eine Sammlung nützlicher Hilfsprogramme zum Aufräumen der Festplatte, die unter einer gemeinsamen grafischen Oberfläche erreichbar sind.

Dabei steht FS für FileSystem und lint für Staub.

Mit ihnen lassen sich u.a. folgende Dinge auf der Festplatte aufspüren:

- doppelte Dateien (nützlich z.B. bei großen Musiksammlungen)
- leere Verzeichnisse
- Dateien mit nicht UTF8-konformen Namen
- temporäre Dateien
- fehlerhafte symbolische Verknüpfungen
- ungültige IDs in der Benutzer und Gruppenverwaltung

# Dublettensuche

## FSlint

FSlint ist eine Sammlung nützlicher Hilfsprogramme zum Aufräumen der Festplatte, die unter einer gemeinsamen grafischen Oberfläche erreichbar sind.

Dabei steht FS für FileSystem und lint für Staub.

Mit ihnen lassen sich u.a. folgende Dinge auf der Festplatte aufspüren:

- doppelte Dateien (nützlich z.B. bei großen Musiksammlungen)
- leere Verzeichnisse
- Dateien mit nicht UTF8-konformen Namen
- temporäre Dateien
- fehlerhafte symbolische Verknüpfungen
- ungültige IDs in der Benutzer und Gruppenverwaltung
- Dateien mit Namenskonflikten (gleiche Dateinamen, Aliase etc.)

# Dublettensuche

## FSlint

FSlint ist eine Sammlung nützlicher Hilfsprogramme zum Aufräumen der Festplatte, die unter einer gemeinsamen grafischen Oberfläche erreichbar sind.

Dabei steht FS für FileSystem und lint für Staub.

Mit ihnen lassen sich u.a. folgende Dinge auf der Festplatte aufspüren:

- doppelte Dateien (nützlich z.B. bei großen Musiksammlungen)
- leere Verzeichnisse
- Dateien mit nicht UTF8-konformen Namen
- temporäre Dateien
- fehlerhafte symbolische Verknüpfungen
- ungültige IDs in der Benutzer und Gruppenverwaltung
- Dateien mit Namenskonflikten (gleiche Dateinamen, Aliase etc.)

# Dublettensuche

## FSlint

FSlint ist eine Sammlung nützlicher Hilfsprogramme zum Aufräumen der Festplatte, die unter einer gemeinsamen grafischen Oberfläche erreichbar sind.

Dabei steht FS für FileSystem und lint für Staub.

Mit ihnen lassen sich u.a. folgende Dinge auf der Festplatte aufspüren:

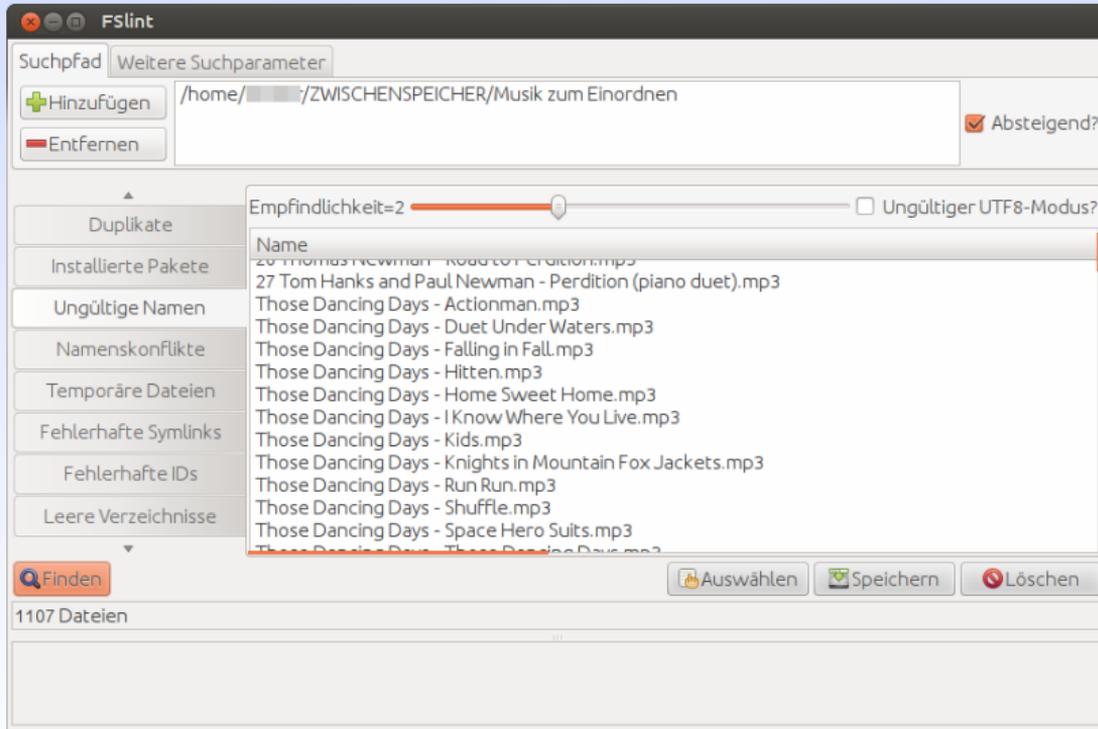
- doppelte Dateien (nützlich z.B. bei großen Musiksammlungen)
- leere Verzeichnisse
- Dateien mit nicht UTF8-konformen Namen
- temporäre Dateien
- fehlerhafte symbolische Verknüpfungen
- ungültige IDs in der Benutzer und Gruppenverwaltung
- Dateien mit Namenskonflikten (gleiche Dateinamen, Aliase etc.)

Nach der Installation kann man das Programm über den Befehl `fslint-gui` ausführen.

`fslint` hat eine intuitiv bedienbare deutsche Oberfläche, mit der man gefundene Dubletten und Dateileichen auch sofort löschen kann.

# Dublettensuche FSlint

# Dublettensuche FSlint



# Dublettensuche

## DupeGuru

# Dublettensuche

## DupeGuru

Die Suche nach Duplikaten kann auf Dateinamen oder dem Inhalt basieren. Außerdem können die Ergebnisse mittels einfachen Filter oder auch RegExp eingeschränkt werden.

# Dublettensuche

## DupeGuru

Die Suche nach Duplikaten kann auf Dateinamen oder dem Inhalt basieren. Außerdem können die Ergebnisse mittels einfachen Filter oder auch RegExp eingeschränkt werden.

Die erkannten Duplikate werden auf Wunsch gelöscht, verschoben oder durch Hardlinks ersetzt (bei letzteren bleibt die Verzeichnisstruktur erhalten, der Speicherplatz wird jedoch freigegeben)

# Dublettensuche

## DupeGuru

Die Suche nach Duplikaten kann auf Dateinamen oder dem Inhalt basieren. Außerdem können die Ergebnisse mittels einfachen Filter oder auch RegExp eingeschränkt werden.

Die erkannten Duplikate werden auf Wunsch gelöscht, verschoben oder durch Hardlinks ersetzt (bei letzteren bleibt die Verzeichnisstruktur erhalten, der Speicherplatz wird jedoch freigegeben)

Daneben existierten zwei weitere Varianten des Programms, spezialisiert auf den jeweiligen Einsatzbereich:

# Dublettensuche

## DupeGuru

Die Suche nach Duplikaten kann auf Dateinamen oder dem Inhalt basieren. Außerdem können die Ergebnisse mittels einfachen Filter oder auch RegExp eingeschränkt werden.

Die erkannten Duplikate werden auf Wunsch gelöscht, verschoben oder durch Hardlinks ersetzt (bei letzteren bleibt die Verzeichnisstruktur erhalten, der Speicherplatz wird jedoch freigegeben)

Daneben existierten zwei weitere Varianten des Programms, spezialisiert auf den jeweiligen Einsatzbereich:

dupeGuru Music Edition für Audiodateien und dupeGuru Picture Edition für Bilder.

# Dublettensuche

## DupeGuru

Die Suche nach Duplikaten kann auf Dateinamen oder dem Inhalt basieren. Außerdem können die Ergebnisse mittels einfachen Filter oder auch RegExp eingeschränkt werden.

Die erkannten Duplikate werden auf Wunsch gelöscht, verschoben oder durch Hardlinks ersetzt (bei letzteren bleibt die Verzeichnisstruktur erhalten, der Speicherplatz wird jedoch freigegeben)

Daneben existierten zwei weitere Varianten des Programms, spezialisiert auf den jeweiligen Einsatzbereich:

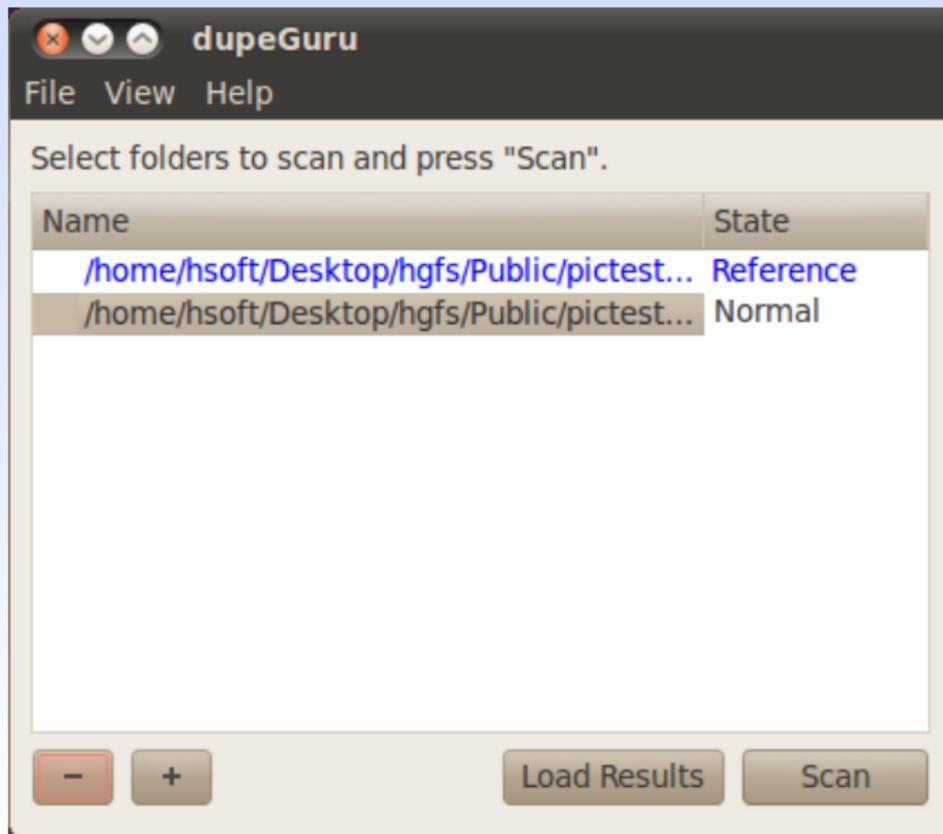
dupeGuru Music Edition für Audiodateien und dupeGuru Picture Edition für Bilder.

Diese spezialisierten Fassungen sind inzwischen Bestandteil des Hauptprogramms, finden sich aber noch in den Paketnamen wieder.

# Dublettensuche

## DupeGuru

# Dublettensuche DupeGuru

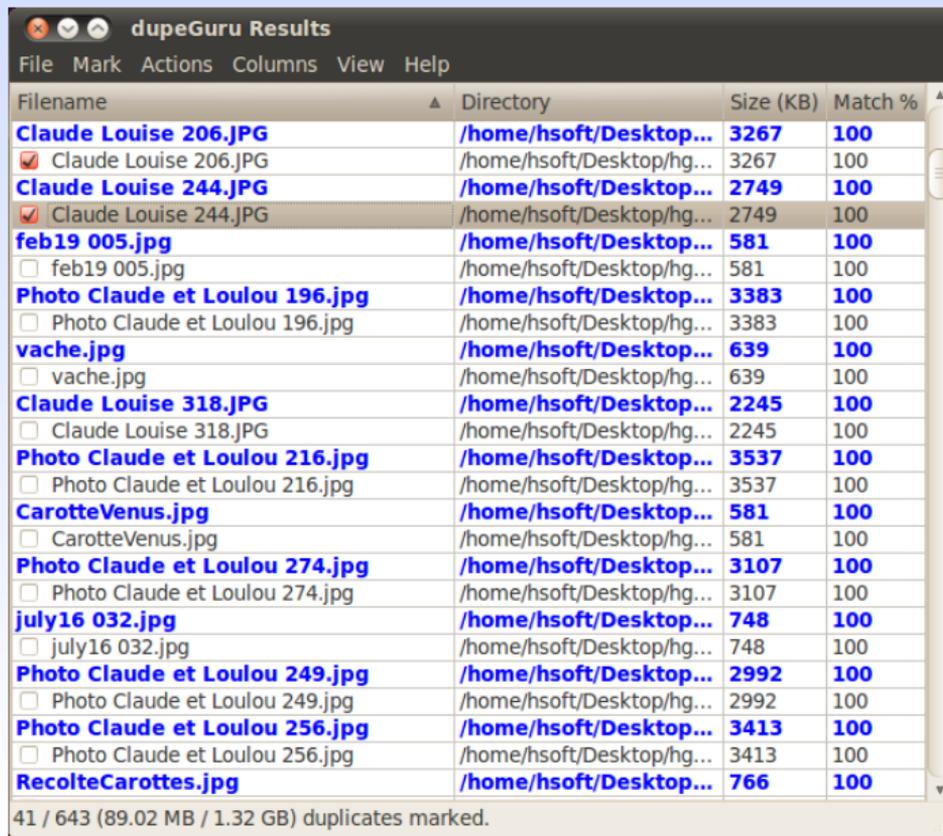


# Dublettensuche

## DupeGuru

# Dublettensuche

## DupeGuru



dupeGuru Results

File Mark Actions Columns View Help

Filename	Directory	Size (KB)	Match %
<b>Claude Louise 206.JPG</b>	/home/hsoft/Desktop/...	3267	100
<input checked="" type="checkbox"/> Claude Louise 206.JPG	/home/hsoft/Desktop/hg...	3267	100
<b>Claude Louise 244.JPG</b>	/home/hsoft/Desktop/...	2749	100
<input checked="" type="checkbox"/> Claude Louise 244.JPG	/home/hsoft/Desktop/hg...	2749	100
<b>feb19 005.jpg</b>	/home/hsoft/Desktop/...	581	100
<input type="checkbox"/> feb19 005.jpg	/home/hsoft/Desktop/hg...	581	100
<b>Photo Claude et Loulou 196.jpg</b>	/home/hsoft/Desktop/...	3383	100
<input type="checkbox"/> Photo Claude et Loulou 196.jpg	/home/hsoft/Desktop/hg...	3383	100
<b>vache.jpg</b>	/home/hsoft/Desktop/...	639	100
<input type="checkbox"/> vache.jpg	/home/hsoft/Desktop/hg...	639	100
<b>Claude Louise 318.JPG</b>	/home/hsoft/Desktop/...	2245	100
<input type="checkbox"/> Claude Louise 318.JPG	/home/hsoft/Desktop/hg...	2245	100
<b>Photo Claude et Loulou 216.jpg</b>	/home/hsoft/Desktop/...	3537	100
<input type="checkbox"/> Photo Claude et Loulou 216.jpg	/home/hsoft/Desktop/hg...	3537	100
<b>CarotteVenus.jpg</b>	/home/hsoft/Desktop/...	581	100
<input type="checkbox"/> CarotteVenus.jpg	/home/hsoft/Desktop/hg...	581	100
<b>Photo Claude et Loulou 274.jpg</b>	/home/hsoft/Desktop/...	3107	100
<input type="checkbox"/> Photo Claude et Loulou 274.jpg	/home/hsoft/Desktop/hg...	3107	100
<b>july16 032.jpg</b>	/home/hsoft/Desktop/...	748	100
<input type="checkbox"/> july16 032.jpg	/home/hsoft/Desktop/hg...	748	100
<b>Photo Claude et Loulou 249.jpg</b>	/home/hsoft/Desktop/...	2992	100
<input type="checkbox"/> Photo Claude et Loulou 249.jpg	/home/hsoft/Desktop/hg...	2992	100
<b>Photo Claude et Loulou 256.jpg</b>	/home/hsoft/Desktop/...	3413	100
<input type="checkbox"/> Photo Claude et Loulou 256.jpg	/home/hsoft/Desktop/hg...	3413	100
<b>RecolteCarottes.jpg</b>	/home/hsoft/Desktop/...	766	100

41 / 643 (89.02 MB / 1.32 GB) duplicates marked.

# Dublettensuche

## fdupes

# Dublettensuche

## fdupes

Es wäre nicht GNU/Linux, wenn es mit fdupes nicht auch ein Kommandozeilen-Werkzeug zur Dublettensuche gäbe...

# Dublettensuche

## fdupes

Es wäre nicht GNU/Linux, wenn es mit fdupes nicht auch ein Kommandozeilen-Werkzeug zur Dublettensuche gäbe...

`fdupes -r /home` sucht beispielsweise rekursiv im Home-Verzeichnis nach Dubletten und zeigt diese an.

# Dublettensuche

## fdupes

Es wäre nicht GNU/Linux, wenn es mit fdupes nicht auch ein Kommandozeilen-Werkzeug zur Dublettensuche gäbe...

`fdupes -r /home` sucht beispielsweise rekursiv im Home-Verzeichnis nach Dubletten und zeigt diese an.

`fdupes -r -d /home` zeigt alle Dubletten an und lässt dem User die Wahl, welche gelöscht werden sollen (einzeln bzw. als Bereich).

# Dublettensuche

## fdupes

Es wäre nicht GNU/Linux, wenn es mit fdupes nicht auch ein Kommandozeilen-Werkzeug zur Dublettensuche gäbe...

`fdupes -r /home` sucht beispielsweise rekursiv im Home-Verzeichnis nach Dubletten und zeigt diese an.

`fdupes -r -d /home` zeigt alle Dubletten an und lässt dem User die Wahl, welche gelöscht werden sollen (einzeln bzw. als Bereich).

Für eine rigorose Löschung ohne Nachfrage kann man die `noprompt (-N)` Option verwenden

# Dublettensuche

## fdupes

# Dublettensuche

## fdupes

```
ambarish@Ubuntu1704:~/playground$ fdupes /home/ambarish/playground
/home/ambarish/playground/file2.txt
/home/ambarish/playground/file3.txt
/home/ambarish/playground/file4.txt
/home/ambarish/playground/file5.txt
/home/ambarish/playground/file1.txt

ambarish@Ubuntu1704:~/playground$ fdupes -d /home/ambarish/playground
[1] /home/ambarish/playground/file2.txt
[2] /home/ambarish/playground/file3.txt
[3] /home/ambarish/playground/file4.txt
[4] /home/ambarish/playground/file5.txt
[5] /home/ambarish/playground/file1.txt

Set 1 of 1, preserve files [1 - 5, all]: 1-5, 1

[+] /home/ambarish/playground/file2.txt
[-] /home/ambarish/playground/file3.txt
[-] /home/ambarish/playground/file4.txt
[-] /home/ambarish/playground/file5.txt
[-] /home/ambarish/playground/file1.txt

ambarish@Ubuntu1704:~/playground$
```

# Dublettensuche

## Ein Wort der Warnung

# Dublettensuche

## Ein Wort der Warnung

Man sollte unter GNU/Linux (vor Allem in der Konsole) immer in etwa wissen, was man tut!

# Dublettsuche

## Ein Wort der Warnung

Man sollte unter GNU/Linux (vor Allem in der Konsole) immer in etwa wissen, was man tut!

Wenn man ein neues (Kommandozeilen-)Werkzeug ausprobiert, bietet es sich an, dieses in einer „sicheren“ Umgebung, z.B. einer eigens eingerichteten Test-Verzeichnisstruktur auszuprobieren.

# Dublettsuche

## Ein Wort der Warnung

Man sollte unter GNU/Linux (vor Allem in der Konsole) immer in etwa wissen, was man tut!

Wenn man ein neues (Kommandozeilen-)Werkzeug ausprobiert, bietet es sich an, dieses in einer „sicheren“ Umgebung, z.B. einer eigens eingerichteten Test-Verzeichnisstruktur auszuprobieren.

Auf diese Weise kann die genaue Funktionsweise getestet werden, bevor man es auf ein echtes Verzeichnis loslässt...

# Links zur Präsentation

<https://wiki.ubuntuusers.de/fslint/>

<https://dupeguru.voltaicideas.net/>

<https://wiki.ubuntuusers.de/DupeGuru/>

<https://github.com/adrianlopezroche/fdupes>

<https://en.wikipedia.org/wiki/Fdupes>

# Weitere Informationen bekommen Sie hier:

`http://www.FreieSoftwareOG.org`  
und  
`Kontakt@FreieSoftwareOG.org`

oder kommen Sie doch einfach zu unserem regelmäßigen Treffen,  
jeden 1. Mittwoch im Monat ab 20:00 Uhr.  
(Treffpunkt und Thema laut Webseite)

