



Troisdorf, 11. Juni 2025

## Einleitung

Der *Filecrawler* ist ein effizientes Werkzeug zum Suchen von Dateien und Ordnern auf modernen Windows PCs. Er wurde in einer virtuellen Windows 7 Maschine mit Delphi XE2 entwickelt und läuft auch unter Windows 10 und wohl auch Windows 11. Es handelt sich um eine 64 Bit Anwendung, die nur aus der „FileCrawler.exe“ Datei und diesem Benutzerhandbuch besteht. Er ist also portabel. Es werden keine Daten mit dem Internet ausgetauscht.

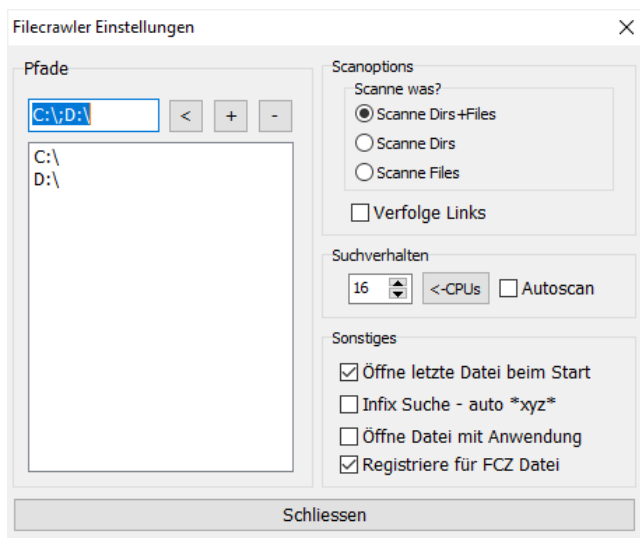
Der *Filecrawler* arbeitet dabei parallel und nutzt die bestehenden Rechenkerne der CPUs. Somit lassen sich selbst Listen mit mehreren Millionen Dateien mittels Live-Suche<sup>1</sup> finden. Als Anzeige der Ergebnislisten dient der sehr schnelle [Virtual-TreeView](#) von Mike Lischke und Lars Dybdahl, der unter der Mozilla Public License 1.1 lizenziert ist.

Der *Filecrawler* behandelt die eingelesenen Dateien nur lesend, speichert seine Einstellungen aber in der „Filecrawler.ini“ Datei im Dokumentenordner und kann Suchergebnislisten oder gesamte Scans der Verzeichnisse in sogenannten FCZ-Dateien speichern. Dies sind komprimierte Dateien, die bei rund einer Millionen Dateien rund 25 MB beanspruchen.

Ich hatte schon den Effekt, dass der *FileCrawler* zwar startet, aber das Hauptfenster nicht erscheint. In diesem Fall scheint irgendwas mit der FileCrawler.ini Datei im Dokumentenordner des Benutzers nicht in Ordnung zu sein. Man kann die Datei dann ruhig löschen. Die Einstellungen gehen dann allerdings verloren, aber der *FileCrawler* sollte wieder starten.

## Der erste Start

Nach dem ersten Start sollte man mittels „Datei → Einstellungen“ in folgenden Dialog wechseln:



Die Pfade bezeichnen die Wurzelverzeichnisse eines PCs (oder aber auch UNC-Pfade) die mit „+“ der Liste hinzugefügt werden oder durch selektieren eines Listeneintrags und dann mit „-“ entfernt werden können. Der Button „<“ liest die Pfade des Systems aus und schreibt ins links stehende Feld und können anschließend mit „+“ hinzugefügt werden.

Oben rechts mit „Scanne was?“ steht, was gescannt werden soll. Ich empfehle „Dirs + Files“ zu scannen, da man Verzeichnisse mit „\*\“ filtern kann.

---

<sup>1</sup> Die Ergebnisse kommen beim Tippen

„Verfolge Links“ sollte, so meine Empfehlung, deaktiviert sein, damit keine Junctions oder SymLinks verfolgt werden. Dies kann in Zyklen münden, deren Erkennung recht schwierig ist und auch Werte wie Dateigrößen doppelt zählt etc..

Beim „Suchverhalten“ wird festgelegt mit wie vielen Kernen die Listen durchsucht werden. Wenn man z.B. 1.000.000 Dateien hat, dann wird die Suche bei z.B. 8 Kernen nur jeweils auf 125.000 Elementen durchgeführt. Der „CPUs“ Knopf liest die Kernanzahl des Systems aus und übernimmt sie.

Die „Autoscan“ Option gibt an, ob der Scan über die Verzeichnisse schon zum Programmstart ausgeführt werden soll.

Unter „Sonstige“ befinden sich die Möglichkeit die zuletzt geladene FCZ-Datei auch schon beim Start zu laden.

Die „Infix-Suche“ baut um einen Suchstring das \* davor und \* danach. Das heißt es werden dann alle Dateien/Verzeichnisse gesucht, die den Text enthalten. Dies hat aber den Nachteil, dass man keine Verzeichnisse mit z.B. \*\ finden kann. Für fortgeschrittene Anwender empfehle ich die „Infix-Suche“ nicht.

Der *Filecrawler* bietet die Möglichkeit per **Doppelklick** auf ein Ergebnis die Datei im Explorer oder eben mit der verknüpften Anwendung zu öffnen. Dies regelt die Einstellung „Öffne Datei mit Anwendung“.

Registrierte für *FCZ-Datei* legt fest, ob der *Filecrawler* beim klicken auf eine *FCZ-Datei* geöffnet werden soll oder nicht. Dazu wird nur auf die Registrierung des Benutzers zugegriffen (HKCU). Falls es Probleme geben sollte, sollte die Checkbox deaktiviert und aktiviert werden. Die Registry-Einträge werden dann neu geschrieben bzw. entfernt.

Damit sind die Einstellungen abgeschlossen und müssen in der Regel auch nicht mehr verändert werden.

Widmen wir uns nun den einzelnen Tabs der Anwendung, nachdem kurz auf das Suchen eingegangen wurde.

## Suchen

Sofern die Infix-Suche deaktiviert ist (empfohlen für versierte Anwender), lassen sich Dateien/Verzeichnisse mit Wildcards aufspüren. Diese sind \* und ?. Dabei steht der \* für eine beliebige oder keine Zeichenfolge und das ? für ein beliebiges Zeichen. Weiterhin lässt sich mit | eine ODER Semantik realisieren. \*.jpg|\*.gif listet z.B. alle JPG und GIF Dateien auf. Die ODER Semantik greift nicht bei der Typanalyse und bei den Dubletten.

Wenn die Infix-Suche aktiviert wird, geht ODER nicht, es wird dann ein \*SUCHSTRING\* gebildet.



## Quelldateien

Im Hauptfenster lassen sich mit „StartScan“ die Einleseprozesse starten. Dabei wird für jeden Eintrag der Liste der Pfade aus den Einstellungen ein separater Scanner-Thread gestartet. Die Wurzeln dürfen auch UNC-Pfade sein, was der Analyse im Netzwerk dienlich sein sollte.

Wenn alle Datei eingelesen sind, kann man mit der Suche mit den beschriebenen Suchmustern beginnen. Dabei handelt es sich um eine Live-Suche, die angestoßen wird, sobald sich der Suchstring ändert.

Über das Menü „Datei“ lassen sich alle gefundenen Dateien/Verzeichnisse Speichern („Datei → Speichern“, „Datei → Speichern unter...“) oder das Ergebnis einer Suche („Auswahl speichern“, „Auswahl speichern unter...“). Der Dateityp ist eine FCZ Datei, die gut komprimiert ist.

Die Ergebnislisten zeigen auch die Icons für im System verknüpfte Dateien an, die beim Anzeigen der Ergebnisse per Cache ermittelt werden. Dabei wird auch überprüft, ob eine Datei oder ein Verzeichnis existiert. Sollte das NICHT der Fall sein, kommen folgende Symbole zum Einsatz:

-  - Die Datei existiert nicht mehr,
-  - Der Ordner existiert nicht mehr.

**Bei allen Ergebnislisten, lassen sich die Spalten durch klicken auf die Überschriften auf- oder absteigend sortieren!**

## Geladen

Dieser Tab enthält die Ergebnisse, die aus einer geladenen *FCZ-Datei* stammen. Dies können ganze Scans sein oder eben auch Suchergebnisse, die mit „Auswahl speichern“, gespeichert wurde. Die Suche verhält sich hier natürlich identisch, wie schon unter *Quelldateien*. Sollte das automatische Laden in den Einstellungen aktiviert sein, werden hier die Ergebnisse nach dem Programmstart bereitgestellt.

## Typanalyse

Die *Typanalyse* sieht die Möglichkeit vor, die Dateien nach ihrer Endung zu untersuchen. Dafür gibt es die zwei Schaltflächen „Analysiere Quelle“ und „Analysiere Geladen“. Je nach Wahl, werden alle Typen basierend auf dem Scan oder der geladenen Datei analysiert und in der Liste dargestellt. Die Liste lässt sich auch nach Endungen durch das Suchfeld oben links filtern. Ein Klick auf einen Dateityp liefert alle Dateien dieses Typs im rechten *Virtual-TreeView*. Die ODER-Semantik wurde hier aber nicht implementiert, da die Suche nur auf Dateiendungen beschränkt ist.

## Dubletten






Die Suche nach mehrfach vorhandenen Dateien ist ein AddOn, was in Grundzügen ganz gut funktioniert. Hierzu finden sich wie bei der Typanalyse wieder die Knöpfe für die Datenquelle. Die Dateien werden dann nach Größe verglichen und je nachdem, ob zwei Dateien/Verzeichnisse oder mehr die gleiche Größe haben, werden sie in eine Liste potentiell gleicher Dateien/Verzeichnisse übernommen. Durch Klick auf die linke Liste werden rechts die potentiell identischen Dateien angezeigt.

Die Dateien sind nach Anzahl der Dateien oder Größe je nach Wahl sortiert. Im Suchfeld oben links lassen sich Suchmuster eingeben. Sollte in einer Gruppe möglich gleicher Dateien eine sein, die dem Suchmuster entspricht, wird die Gruppe weiterhin angezeigt.

Die Schaltfläche „Dateivergleich“ startet dann ein Vergleichs-Thread der die potentiell identischen Kandidaten auf Dateiebene bestimmt. Hierbei ist zu beachten, dass wenn man z.B. 100 potentiell gleiche Dateien hat, jede mit jeder verglichen werden muss, um sie in Gruppen identischer Dateien einzuordnen. In der Statusbar finden sich dann Angaben zum Fortschritt. Sollten alle Dateien unterschiedlich sein, dann wird der Eintrag aus der Liste links gelöscht. Wenn identische Dateien gefunden werden, dann erscheinen sie rechts unterhalb der Gruppe.

Das Menü Dubletten bietet noch Einstellmöglichkeiten, ob Verschachtlung geprüft werden soll oder ob Verzeichnisse mit einbezogen werden sollen. Verzeichnisse identischer Größe werden noch nicht ausführlich behandelt. Interessant ist aber der „Autovergleich“: Damit wird der Dateivergleich angestoßen, sobald sich im linken VST der Fokus verschiebt.

Für die Darstellung der Dubletten im rechten Teil der Anwendung kommen folgende Icons zur Anwendung:

-  - Die Datei ist gleich groß wie die anderen, aber nicht notwendigerweise identisch.
-  - Der Ordner hat dieselbe Größe wie die anderen Einträgen und ist möglicherweise identisch.
-  - Die Datei wurde mit den anderen der Gruppe Byte für Byte verglichen und ist identisch.
-  - Dies wurde noch nicht implementiert, soll aber die Gleichheit ganzer Ordner anzeigen.
-  - Dies ist das Symbol für eine Gruppe identischer Dateien.

## Hinzugefügte

Beim Ergebnis dieses Tabs macht der *FileCrawler* eine Mengenberechnung, welches für jede Quelldatei/Verzeichnis überprüft, ob sie in der Liste der geladenen Dateien enthalten ist. Sollte dies der Fall sein, wird die Datei NICHT übernommen. Es handelt sich also um die Mengenfunktion **Quelle MINUS geladene Datei**. Das ergibt die hinzugefügten Dateien.

## Gelöschte

Dies ist die umgekehrte MINUS Funktion, **geladene Dateien MINUS Quelle**. Daraus erscheinen dann die nicht mehr verbliebenen, also gelöschten Dateien.

## Schlusswort

Ich bedanke mich für die Aufmerksamkeit. Für Rückfragen und Anmerkungen bin ich gerne unter der Email Adresse [filecrawler@np-doc.de](mailto:filecrawler@np-doc.de) erreichbar.

Ich hoffe, die Anwendung gefällt Ihnen.

Let's Crawl!

Mit freundlichen Grüßen  
Ulrich Daus

## **Lizenz**

Die Lizenz ist auch in der Applikation hinterlegt. Der Vollständigkeit nochmal hier:

*FileCrawler – Freeware License  
Copyright (c) 2025 Ulrich Daus*

*Dieses Programm ist Freeware. Es darf kostenlos genutzt, kopiert und unverändert weitergegeben werden, solange diese Lizenzdatei enthalten bleibt.*

*Nutzung:*

- *Die Software darf für private und berufliche Zwecke verwendet werden.*
- *Eine kommerzielle Weiterverbreitung oder der Verkauf ist nicht gestattet ohne vorherige Zustimmung des Autors.*
- *Änderungen am Programm oder an den Begleitdateien sind nicht erlaubt.*

*Haftungsausschluss:*

*Diese Software wird "wie sie ist" bereitgestellt, ohne jegliche Garantie. Der Autor haftet nicht für Schäden, die direkt oder indirekt durch die Nutzung dieser Software entstehen.*

*Kontakt:*

*filecrawler@np-doc.de*

*Stand: Juni 2025*