

FileCrawler

Benutzerhandbuch
Ab Version 1.2.0 Beta



Troisdorf, 24. Oktober 2025

Neu:
Grundlegende Überarbeitung der Mengenfunktionen (siehe Ende)

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 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-Suche¹ 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.

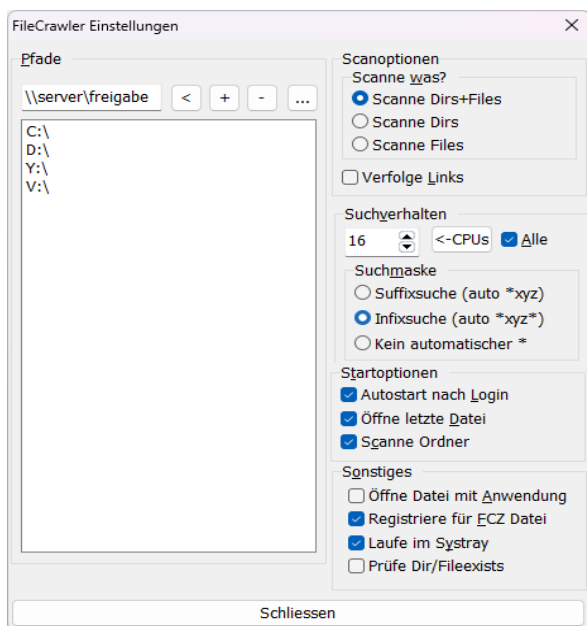
Der *Filecrawler* behandelt die eingelesenen Dateien nur lesend, speichert seine Einstellungen aber in der „Filecrawler.ini“ Datei im Dokumentenordner des Users oder bevorzugt wegen der Portabilität im Anwendungsordner. Der *Filecrawler* kann Suchergebnislisten oder gesamte Scans der Verzeichnisse in sogenannten FCZ-Dateien speichern. Dies sind komprimierte Dateien, die bei rund einer Millionen Dateien rund 25 MB beanspruchen. Somit lassen sich auch die Verzeichnisse von externen Speichermedien wie USB-Sticks offline durchsuchen!

Der erste Start

Wenn keine *FileCrawler.ini* Datei vorhanden ist, dann wird das als erster Start des *FileCrawlers* interpretiert. In diesem Fall werden **ab Version 1.1.2 Beta** die Quellpfade des Systems ausgelesen und in die Liste eingetragen. Auch die Anzahl der CPU Kerne wird ausgelesen und übernommen.

Um die Werte zu überprüfen, gelangt man über „Datei → Einstellungen“ in folgenden Dialog:

¹ Die Ergebnisse kommen beim Tippen



Die Pfade bezeichnen die Wurzelverzeichnisse, die rekursiv eingelesen werden. Dabei wird für jeden Pfad der Liste der Reihe nach ein Scan-Prozess gestartet. Es sind auch UNC-Pfade (im Format \\Server\Freigabe) möglich.

Der Button „<“ liest die Pfade des Systems aus und trägt sie in die Liste ein. Mit „+“ werden die durch „;“ getrennten Pfade der Liste hinzugefügt. Mit „-“ wird ein selektierter Pfad entfernt. Mit „...“ kann ein Ordner gewählt werden, der dann angefügt wird.

Die Reihenfolge der Pfade lässt sich per Drag&Drop bestimmen.

Sollte nicht mit dem System verbundene Pfade in der Liste sein, so macht das nichts.

In den „Scanoptionen“ oben rechts steht mit „Scanne was?“, was gescannt werden soll: Dateien&Verzeichnisse, nur Verzeichnisse oder nur Dateien. Ich empfehle „Dirs + Files“ zu scannen, da man Verzeichnisse mit „*\“ filtern kann.

„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 CPU-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. Es wird empfohlen alle logischen Kerne zu benutzen. Die Einstellung „Alle“ ist standardmäßig aktiviert und bewirkt, dass bei jedem Programmstart die maximale Anzahl an Kernen verwendet wird. Dies ist interessant, wenn man portabel unterwegs ist.

Im Auswahlfenster „Suchmaske“, wird festgelegt ob Suchstrings automatisch mit einem * beginnen (Suffixsuche), ob sie automatisch mit * beginnen und * enden (Infixsuche) oder ob man selber die * eingibt. Nähere Informationen finden sich im Abschnitt *Suchen*.

Neu in **Version 1.1.3 Beta** sind die Startoptionen:

Der *FileCrawler* kann auf Wunsch nach dem Login starten („Autostart nach Login“). Darunter wird angegeben ob die zuletzt geladene Datei beim Start des *FileCrawlers* geladen werden soll („Öffne letzte Datei“)². Ferner kann angegeben werden, ob die hinterlegten Ordner beim Start gescannt werden sollen („Scanne Ordner“).

Der *Filecrawler* bietet die Möglichkeit per **Doppelklick** auf ein Ergebnis die Datei oder das Verzeichnis im Explorer zu öffnen. Dateien lassen sich aber auch direkt mit der verknüpften Anwendung öffnen. Dies regelt die Einstellung „Öffne Datei mit Anwendung“.

² Dies geschieht nun in einem eigenen Prozess.

„Registrierte für FCZ-Datei“ legt fest, ob der *FileCrawler* beim Klicken auf eine *FCZ-Datei* im Windows Explorer 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.

Neu hinzugekommen ist in der **Version 1.1.3 Beta** die Option „Laufe im Systray“. Dies ist der Bereich bei der Uhr in der Startleiste. Damit läuft der *FileCrawler* nachdem man das Hauptfenster per Klick auf's **X** geschlossen hat im Hintergrund und lässt sich leicht reaktivieren. Wenn zusätzlich die Autostart Option gewählt wurde, läuft der *FileCrawler* nach der Anmeldung minimiert im Systray. Beenden lässt sich der *FileCrawler* dann über das Menü „Datei → Beenden“ oder „ALT+F4“ (sofern das Hauptfenster aktiv ist), sowie über das Kontextmenü über dem Systray Symbol.

Neu in **Version 1.2.0 Beta** ist die Option „Prüfe Dir/Fileexists“. Diese Einstellung regelt, ob bei der Anzeige von Einträgen die Existenz von Dateien oder Verzeichnissen geprüft wird. Bei älteren HDDs (vor allem, wenn sie in den Ruhemodus gehen) kann das Einschalten zu Verzögerungen führen. Bei SSDs hingegen ist die Prüfung problemlos. (Siehe auch Tab Quelldateien.)

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

Widmen wir uns nun zunächst der Suche und den Menüs, um dann abschließend die einzelnen Tabs der Anwendung zu beschreiben.

Suchen

Die Suchstrings in dieser Anwendung können die Wildcards ***** und **?** enthalten. Dabei steht der ***** für eine beliebige oder keine Zeichenfolge und das **?** für ein beliebiges Zeichen. Ferner gibt es das ODER Zeichen **|** über das eine Oder-Semantik realisiert wird. Sofern die Suchmaske leer ist oder nur aus ***** besteht, werden alle Einträge gefunden. Die Suche ist NICHT Case-Sensitive, d.h. Groß- und Kleinschreibung ist egal. Doppelte ****** werden zu einem ***** zusammengefasst.

In den Einstellungen der Suchmaske lässt sich festlegen, ob ***** automatisch für eine bequemere Suche gesetzt werden. Dies ist bei der Suffixsuche (Standardeinstellung) und Infixsuche der Fall. Wenn in ein Suchfeld z.B. *xyz* eingegeben wird, dann macht die Suffixsuche ein **xyz* daraus und sucht alle Einträge die auf *xyz* enden. Die Infixsuche macht aus *xyz* ein **xyz**, was bedeutet dass alle Einträge gefunden werden, die *xyz* enthalten. Bei der manuellen Suche wird der Suchstring *xyz* nichts finden, weil die Laufwerke ja z.B. *C:* und nicht *xyz* heißen. Einen Vorteil hat aber auch die manuelle Suche:

Mittels *c:\windows** lassen sich z.B. alle Dateien/Verzeichnisse unterhalb von *c:\Windows* finden. Mit *c** alle Dateien unterhalb von Laufwerk C. Bei der Suffix- oder Infixsuche würde hingegen nach **c** gesucht werden, was sehr viele Ergebnisse produzieren würde.

Interessant ist auch die Oder Semantik mit **|**. Sucht man bei aktivierter Suffixsuche z.B. nach *.gif|.jpg*, so werden alle Dateien gefunden, die auf **.gif* ODER **.jpg* enden. Wäre die Infixsuche an, so würde nach **.gif** oder **.jpg** gesucht. Die Manuelle Suche würde hierbei nichts finden.

Zur Verdeutlichung der Suche hier noch einige Beispiele:

1. **windows** findet alle Dateien/Ordner, die *windows* als Text enthalten. Bei der Infixsuche braucht man die *** nicht einzugeben.
2. *c*windows** liefert bei manueller Suche alle Dateien/Ordner auf Laufwerk C, die *windows* enthalten.
3. **meinbild.gif* findet alle Dateien/Ordner, die auf *meinbild.gif* enden. Bei der Suffixsuche braucht man auch hier den *** nicht.
4. **meinbild.?if* findet alle Dateien/Ordner, die auf *meinbild.gif* oder aber auch z.B. *meinbild.tif* enden. Anstelle des *?* kann ein beliebiges Zeichen stehen.
5. **.???* findet alle Dateien die eine dreistellige Endung haben.
6. *** findet alle Verzeichnisse - bei der Suffixsuche reicht ein **.
7. ****** Findet vier oder mehr als vierfach verschachtelte Verzeichnisse.
8. **abc??*xyz** findet alle Dateien/Verzeichnisse in denen *abc* vorkommt danach zwei beliebige Zeichen und dann dahinter *xyz*.
9. **.gif|.tif|.bmp* findet alle GIF, TIF oder BMP Dateien

Ich hoffe, dass die Wildcard-Suche jetzt verständlich ist und möchte betonen, dass man die höchste Flexibilität mit der manuellen Suche hat, die ich versierten Anwendern ans Herz lege.

Das Menü Datei

Mittels dem Speichern & Laden von sogenannten FCZ-Dateien, lassen sich Listen von Verzeichnissen und Dateien archivieren. Eine FCZ-Datei ist gut komprimiert.

Damit ist es auch möglich Dateien/Verzeichnisse von externen Speichermedien komprimiert zu speichern und offline, d.h. ohne das das Medium mit dem Computer verbunden ist, zu durchsuchen.³

Über das Menü „Datei“ lassen sich FCZ Dateien öffnen (ab **Version 1.1.3 Beta** als Hintergrund Prozess) oder Dateien auch anfügen, d.h. die Werte einer zuvor geladenen Datei werden behalten. Ferner lassen sich alle gescannten oder geladenen Einträge speichern („Datei → Speichern“, „Datei → Speichern unter...“) oder das Ergebnis einer Suche („Auswahl speichern“, „Auswahl speichern unter...“). Bei der Typanalyse z.B. werden über „Auswahl speichern“ nur die Ergebnisse des rechten Virtual-Treeview gespeichert.

Das Menü Schnellscan...

...ist hinzugekommen, um auch Sticks oder externe Festplatten schnell scannen (und womöglich speichern) zu können. Das Menü wird bei einem Wechsel der Datenträger dynamisch aufgebaut und bietet die Möglichkeit ein Laufwerk neu einzulesen (die bisherige Quelle wird geleert), oder der Quelle hinzuzufügen – dies läuft dann als eigener Prozess.

Ab **Version 1.1.2 Beta** können hier auch einzelne Ordner gescannt oder den Quelldateien hinzugefügt werden.

Das Menü Dubletten...

...wird im Tab: Dubletten beschrieben.

3 Die Erweiterung des *FileCrawler* um einen Stick-Manager oder wie es auch heißen wird, ist geplant.

Das Menü Info...

...bedarf wohl keiner Erklärung.

Kommen wir nun zu den einzelnen Tabs der Anwendung:

Tab: Quelldateien

Im Hauptfenster lassen sich mit dem Knopf „Scanne gespeicherte“ die Einleseprozesse der unter Einstellungen festgelegten Pfade 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. Man kann aber auch mit dem „Schnellscan“ Menü arbeiten und Laufwerke einzeln einlesen.

Wenn alle Dateien/Ordner 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.

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 UND ab **Version 1.2.0 Beta** die Option „Prüfe Dir/Fileexists“ an 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! Die Positionen der Spalten lassen sich ebenfalls verändern. All dies wird in der *FileCrawler.ini* Datei gespeichert.

Tab: 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 wurden. 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.

Tab: 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*.

In dem rechten Suchfeld kann in den Ergebnissen des rechten Virtual-TreeViews gesucht werden. **Dabei ist eine Infixsuche realisiert**, d.h. man braucht keine * einzugeben und erhält Ergebnisse.

Tab: 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 oder Dateien gesucht 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.

Tab: Mengenfunktionen

Dieser Programmabschnitt wurde in **Version 1.2.0 Beta** grundlegend überarbeitet. Die einfachen Berechnungen mit „Hinzugefügte“ und „Gelöschte“ sind dem Tab **Mengenfunktionen** gewichen, der auch die alten Funktionen abbildet. Hier kann man nun auf unterschiedliche Art *Quelle A* und *Quelle B* spezifizieren und dann die Schnittmenge und Differenzmengen berechnen. Die Quellen können

- die gescannten Verzeichnissen (absoluter Vergleich),
- ein Ordner aus den gescannten Verzeichnissen (relativer Vergleich),
- und entsprechend die geladenen Verzeichnisse (absoluter Vergleich),
- ein Ordner der geladenen Verzeichnisse (relativer Vergleich),
- eine komplette FCZ Datei (absoluter Vergleich) oder zu guter Letzt
- ein Ordner der geladenen FCZ Datei (relativer Vergleich)

sein.

Beim **absoluten Vergleich** werden die Pfade samt ihren Laufwerksangaben gegeneinander geprüft. Beim **relativen Vergleich** werden die Dateien und Verzeichnisse unterhalb der angegebenen Ordner für *Quelle A* und *Quelle B* gegeneinander berechnet. Nach der Berechnung werden die relativen Pfade je nach Ergebnis wieder um den Ordnerpfad ergänzt.

Es wird sowohl die Schnittmenge als auch die Differenzmengen *Quelle A MINUS Quelle B* sowie *Quelle B MINUS Quelle A* berechnet. Bei der Schnittmenge (Eintrag kommt in Quelle A UND Quelle B vor) reicht es, wenn der Dateiname identisch ist. Das Ergebnis hat dann die Zugehörigkeit „*Namensgleich*“. Sollte neben dem Dateinamen auch die Größe, das Datum und bei Verzeichnissen die Anzahl der Unterverzeichnisse sowie die Anzahl der Dateien gleich sein, dann ist die Zugehörigkeit „*Pot. Identisch*“.

Die Ergebnisse erscheinen farbig unterlegt je nach Art im unteren VST. Die Spalten lassen sich wie gewohnt sortieren. Zusätzlich lässt sich die mitunter große Menge auch durch die vier Checkboxen **Gleiche**, **Gleicher Name**, **A MINUS B** und **B MINUS A** filtern. Ist das Häkchen gesetzt, werden die zugehörigen Ergebnisse angezeigt oder eben abgewählt.

Das aus der vorigen Version bekannte „Hinzugefügte“ sind dann die Ergebnisse mit der Zugehörigkeit „*In A NICHT in B*“ wobei bei *Quelle A* „Alle Quelldateien“ gewählt ist und *Quelle B* „Alle geladenen“ ist. Das aus der alten Version bekannte „Gelöschte“ sind dann die Ergebnisse mit der Zugehörigkeit „*In B NICHT in A*“.

Schlusswort

Ich bedanke mich für die Aufmerksamkeit. Für Rückfragen und Anmerkungen bin ich gerne unter der Email Adresse 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: Oktober 2025