



Troisdorf, 25. Juni 2025

Einleitung

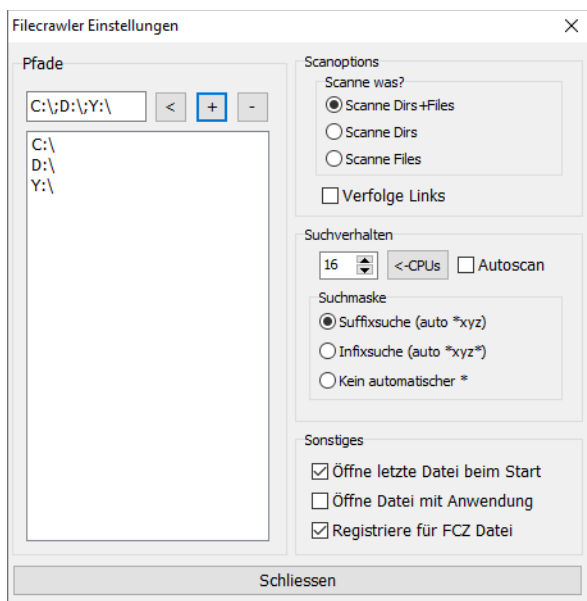
Der *Filecrawler* ist ein effizientes Werkzeug zum Suchen von Dateien und Ordnern auf modernen Windows PCs. Er wurde in einer virtuellen Windows 7 Maschine mit Delphi XE2 entwickelt und läuft auch unter Windows 10 und wohl auch Windows 11. Es handelt sich um eine 64 Bit Anwendung, die nur aus der „FileCrawler.exe“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

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 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.

¹ 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. Es wird empfohlen alle logischen Kerne zu benutzen.

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

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*.

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

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 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

Über das Menü „Datei“ lassen sich alle gescannten oder geladenen Dateien/Verzeichnisse Speichern („Datei → Speichern“, „Datei → Speichern unter...“) oder das Ergebnis einer Suche („Auswahl speichern“, „Auswahl speichern unter...“). Auch lassen sich hier Dateien öffnen, was in den Tab „Geladen“ führt. Der Dateityp ist eine FCZ Datei, die gut komprimiert ist. 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.

Das Menü Dubletten...

...wird im Tab: Dubletten beschrieben.

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, 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.

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: 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.

Tab: Gelöschte

Dies ist die umgekehrte MINUS Funktion, **geladene Dateien MINUS Quelle**. Daraus erscheinen dann die nicht mehr verbliebenen, also gelöschten Dateien. Der Doppelklick auf Ergebnisse wurde hier nicht implementiert.

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: Juni 2025