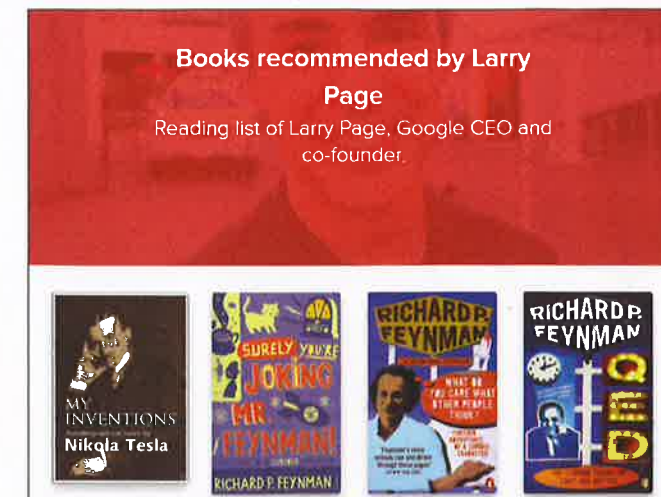


Stoff für den Kopf

www.bookicious.com/collections

<https://www.highlyreco.com>



Mauro Sicard liest gerne. Der 18-jährige Web-Entwickler und „Entrepreneur“ hat aus seinem Hobby ein Projekt gemacht. Er sammelt, welche Bücher Größen aus der IT-Szene wie Bill Gates und Larry Page empfehlen. Diese Kollektion hat er auf der Website **Bookicious** veröffentlicht.

Die englischsprachige Site richtet sich vor allem an Gründer und Maker wie Sicard selbst. Unter den gelisteten Büchern finden sich aber auch Perlen, die jedermann interessieren könnten, zum Beispiel die Einstein-Biographie von Walter Isaacson oder „Moby Dick“. Bookicious enthält die Empfehlungen von 18 IT-Persönlichkeiten, bei der ähnlich gelagerten Site **Highly-reco** sind es sogar 40. (jo@ct.de)

Coder-Mucke

musicforprogramming.net

Beim Programmieren und anderen geistigen Tätigkeiten taucht man am besten in einen Tunnel ab; so kann ein ungemein produktiver Flow entstehen. Am besten, man blendet dafür die ablenkenden Umgebungsgläusche aus oder übertönt sie mit Musik – die idealerweise auch noch die Konzentration fördert. Nach Ansicht des Entwicklers John Davies muss sie dazu zum Beispiel Brummen und Texturen ohne Rhythmus enthalten; Schlagzeug und Gesang dagegen sollten sehr dosiert vorkommen, wenn überhaupt. Nach eigenen Angaben hat er Jahre damit zugebracht, die Musik auf **musicForProgramming** zusammenzutragen. Die Website ist gestaltet wie ein Listing im Editor SublimeText. Man kann die 46 Stücke direkt abspielen oder herunterladen. (jo@ct.de)

Alle Links dieser Seite
unter ct.de/yrvv



Alle Wege führen zur Philosophie

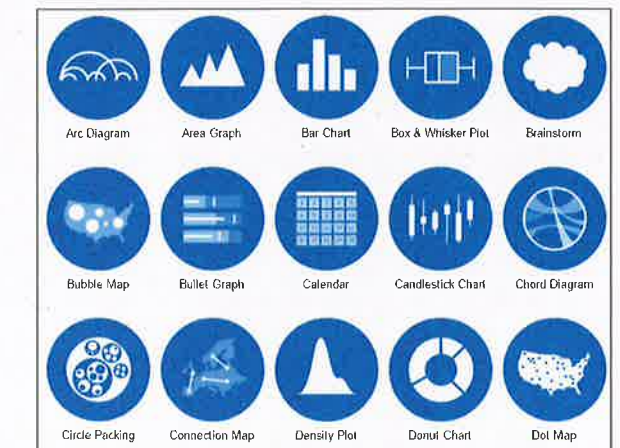
<https://xefer.com/wikipedia>

https://en.wikipedia.org/wiki/Wikipedia:Getting_to_Philosophy

Die Beobachtung ist nicht neu, aber immer wieder faszinierend: Starten Sie bei einem beliebigen Artikel in der Wikipedia, klicken Sie auf den ersten Link und wiederholen sie das für alle weiteren Artikel, zu denen sie im Folgenden kommen. Irgendwann landen sie mit großer Wahrscheinlichkeit beim Eintrag für „Philosophie“. Sie können das aber auch die Website **Xefer** für Sie machen lassen, dort werden die Verknüpfungen hübsch visualisiert. (jo@ct.de)

Daten schön präsentieren

datavizcatalogue.com



Immer nur Tortendiagramme – langweilig, insbesondere weil es für viele Daten geeignetere Visualisierungen gibt. Severino Ribeca präsentiert auf seiner Site **The Data Visualisation Catalogue** 60 Darstellungsmöglichkeiten. Für jede einzelne beschreibt er, welche Typen von Daten sich damit besonders gut verbildlichen lassen. Außerdem stellt er jeweils Alternativen vor sowie Tools, mit denen man vorgestellte Visualisierungen generieren kann. (jo@ct.de)

Hype-Videos

Gruselige Überraschung für Kunden eines Elektroladens: Die fiese Gestalt aus einem **Horrorvideo** – zunächst nur auf einem Fernseher zu sehen – beginnt plötzlich, diesem zu entsteigen.

<https://www.facebook.com/techinsider/videos/663960513802308/> (1:47, Englisch)

Pierre-Alexandre Chauvat hat Ausschnitte aus mehr als 50 Filmen geschickt aneinandergeklebt – heraus kommt ein rasanter Kurzfilm namens **Everybody wants to kill Bruce 2**. <https://vimeo.com/194076406> (12:38, Englisch)

Texterkenner lernt PDF

OCR-Spezialist Abbyy FineReader bearbeitet jetzt auch PDF-Dokumente

Abbyys Texterkennung kann mehr, als Dokumente in bearbeitbare Office-Dateien zu verwandeln: Mit FineReader 14 lassen sich PDFs editieren und Dokumente unterschiedlicher Dateiformate miteinander vergleichen.

Von Anke Poimann und Andrea Trinkwalder

Der Umfang von Abbyy FineReader 14 ist deutlich gewachsen: Zusätzlich zur Texterkennung sind jetzt alle Funktionen der Software PDF Transformer+ Bestandteil des Programms. In der von uns getesteten Corporate-Version lassen sich außerdem Schriftstücke – auch unterschiedlichen Dateiformats – miteinander vergleichen.

Im übersichtlichen Startbildschirm wählt der Anwender zwischen Öffnen, Scannen und Vergleichen. Die moderne Ribbon-Oberfläche zeigt alle wichtigen Funktionen.

Sprachgenie

In Abbyy FineReader verläuft die Zeichenerkennung (OCR, Optical Character Recognition) in drei Schritten: Öffnen des Dokuments, automatische Layout-Analyse, Zeichenerkennung. Eine saubere Vorlage, mit mindestens 300 dpi gescannt, liefert das beste Ergebnis. Schwieriger wird es bei zerknitterten Dokumenten oder solchen mit Flecken. Immer häufiger kommt zudem das Smartphone als Quelle zum Einsatz. Wer für gute Ausleuchtung und ein ruhiges Händchen sorgt, der wird auch mit unterwegs fotografierten Vorlagen verwertbare Ergebnisse erzielen.

Richtig spannend wird es bei leicht verwackelten Schnappschüssen mit verzerrter Perspektive. Hier lässt sich die Erkennungsqualität mit den Werkzeugen zur Vorabbildbearbeitung verbessern. In un-

serem Test haben wir die Infotafel eines Klosters ins Rennen geschickt: Weiße Schrift auf blauem Hintergrund und verzerrte Perspektive brachten im ersten Durchlauf völlig unbrauchbare Ergebnisse. Nachdem wir die Perspektive korrigiert und die Farben invertiert hatten, waren nur noch vergleichsweise wenig Nacharbeiten am Erkennungsergebnis nötig.

Das Programm ermittelt die Sprache der Vorlage selbstständig. Wer sie lieber manuell einstellt, wählt sie aus den 192 verfügbaren Sprachen aus. Unter ihnen befinden sich auch Sonderformen wie Esperanto, Java oder C/C++.

Die Layout-Analyse identifizierte im Test wesentliche Bestandteile wie Textspalten, Überschriften, Tabellen, Listen, Grafiken oder Kopf- und Fußzeilen zuverlässig und setzte sie sauber in Office-Formate um. Sollte es an der ein oder anderen Stelle hapern, bessert man von Hand nach. Irrelevante Bereiche lassen sich von der Erkennung ausschließen.

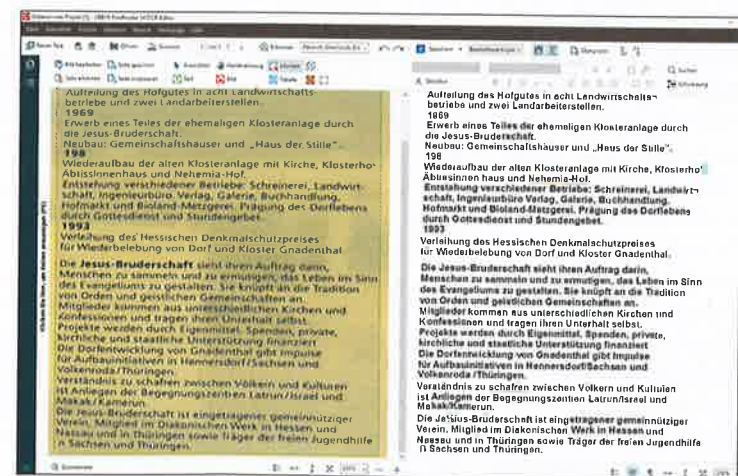
Im Test erkannte Abbyy FineReader sogar einen dreispaltig gesetzten Text mit umflossenen Bildern, ohne Layout und Textfluss zu zerstören. FineReader liegt hier deutlich vor Acrobat, Foxit oder Nuance, insbesondere weil er die deut-

sche Silbentrennung berücksichtigt. Sollte die erkannte Reihenfolge der Absätze nicht stimmen, lässt sich das einfach zurechtrücken: Man klickt im Original die Textblöcke in der gewünschten Reihenfolge an und lässt den Text anschließend neu einlesen.

Das Programm speichert Dokumente entweder in editierbaren Office-Formaten wie DOC(X), XLS(X), RTF und ODT, als PDF und PDF/A sowie in Bild-Formaten wie BMP, JPG, PNG und TIFF. Zusätzlich kann man ein Dokument als EPUB sichern und so sein eigenes E-Book erstellen. Da dieses Format kein festes Seitenlayout kennt, wandelt FineReader 14 erkannte Fußnoten automatisch in verlinkte Endnoten um.

Leseschwierigkeiten

FineReader 12 erkannte lediglich simple chemische Formeln korrekt. Die aktuelle Version soll laut Hersteller auch mathematische Formeln richtig erkennen. Im Test funktionierte das für einzeilige Formeln zuverlässig, Brüche bereiteten dagegen Probleme: An ihrer Stelle erschien eine willkürliche Zeichenfolge, die mit dem Ausgangsmaterial nichts gemeinsam hatte. Sofern man die Formel selbst nicht



Das verzerrte Foto einer Infotafel war zunächst als Vorlage unbrauchbar. Nach der Vorabbildkorrektur ließ sich der Text beinahe fehlerfrei erkennen.

bearbeiten oder später durchsuchen möchte, gibt es einen Ausweg: Man zeichnet im Original an passender Stelle einen Bildbereich ein, sodass es als Bild im erkannten Dokument erscheint.

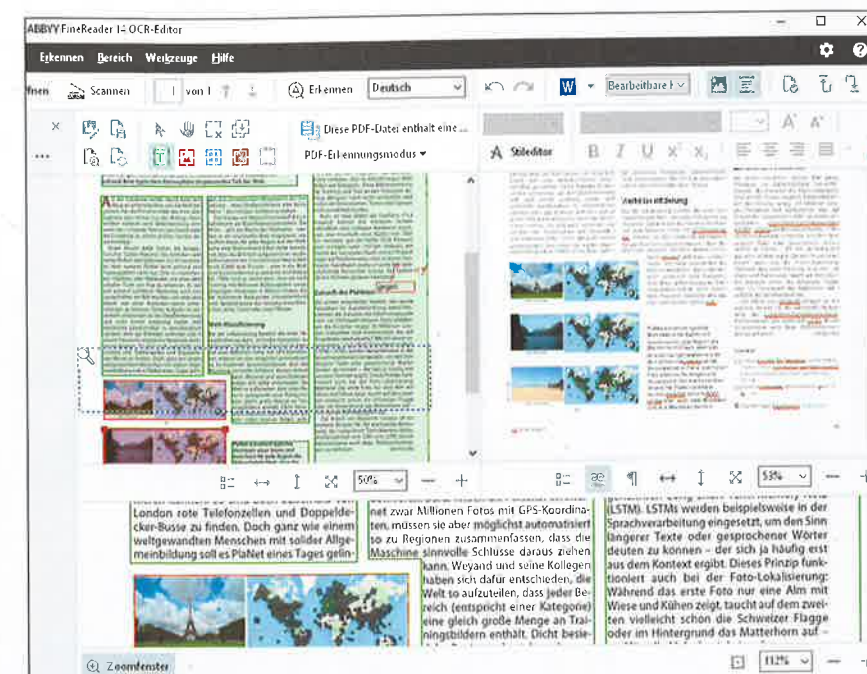
Mit kleinen Schriften hat das Programm keine Schwierigkeiten, mit handschriftlichen Notizen und Fraktur hingegen schon. In unserem Test erzielten wir bis 6 pt korrekte Ergebnisse. Per Textmarker hervorgehobene Passagen in gescannten Dokumenten erkannte FineReader zum Teil nicht korrekt – hier mussten wir nacharbeiten. Alternativ kann man mit der Vorabbildkorrektur Farbmarkierungen auf der Seite entfernen. Das funktioniert jedoch nur bei dunkler Schrift auf hellem Grund. Die Analyse komplizierter Tabellen-Layouts muss mitunter ebenfalls von Hand korrigiert werden.

OCR trifft PDF

Dank der Integration des PDF Transformer+ kann FineReader das Portable Document Format jetzt direkt öffnen und bearbeiten. Außerdem wandelt das Programm PDFs in Office-Formate, exportiert ausgewählte Inhalte, entfernt sensible Daten, hilft beim Ausfüllen von Formularen, bringt Kommentarfunktionen für Review-Prozesse mit und vergleicht Dokumente unabhängig von ihrem Format. Dabei versucht Abbyy die technischen Feinheiten der jeweiligen Formate weitgehend vom Anwender fernzuhalten: FineReader findet Stichwörter nicht nur in PDFs mit Textebene, sondern auch in Nur-Bild-PDFs, deren Seiten gescannt oder abfotografiert wurden. Der Anwender bekommt von dem Unterschied fast nichts mit: Gibt er einen Suchbegriff ein, wirft FineReader bei Bedarf im Hintergrund die Texterkennung an, markiert die Fundstellen im Dokument und listet sie in einer eigenen Leiste sauber mitsamt Kontext auf.

Original und Kopie

Der neu hinzugekommene Dokumentenvergleich ist praxistauglich realisiert. Über das Menü des Windows Explorer oder per Öffnen-Funktion lädt man zwei Dokumente, deren Inhalt FineReader unabhängig vom Format analysiert. Ergebnisse zeigt das Programm in einem Vergleichsfenster an. Sie lassen sich als Word-Datei oder als PDF speichern. Die Word-Datei listet Unterschiede in einer Tabelle auf. Wählt man die PDF-Ausgabe, kann man sich entscheiden, welches der beiden Dokumente als kommentierte Version gespeichert werden soll. So kann man Ori-



Auch PDF-Dokumente mit Spaltensatz und umflossenen Bildern analysiert FineReader per OCR – und macht sie in Office-Anwendungen bearbeitbar.

ginale aus Microsoft Office oder Open Office schnell mit ihren PDF-Pendants vergleichen. Erhält man beispielsweise einen unterschriebenen Vertrag zurück, lässt sich leicht prüfen, ob der Text Änderungen aufweist.

Geht es darum, PDFs in Office-Formate zu konvertieren oder Inhalte zu extrahieren, ist FineReader erste Wahl, weil die Layout-Erkennung PDF-Inhalte ebenso zuverlässig analysiert wie gescannte Dokumente – mitunter sogar ein wenig besser. Das direkte Bearbeiten von PDF-Inhalten beherrscht hingegen Acrobat besser, weil der FineReader-Editor Text nur zeilenweise rekonstruiert, während Acrobat mittlerweile ganze Spalten zu einer Einheit zusammenfasst.

Die Formular-Engine des FineReader ist ausbaufähig. Einfache Formulare mit Textfeldern, Checkboxes, Dropdown-Listen, Berechnungen oder interaktiven Schaltflächen ließen sich im Test problemlos ausfüllen und mitsamt Daten abspeichern. Hinterlegte Skripte ignorierte FineReader dagegen, ohne den Anwender darauf hinzuweisen.

Fazit

FineReader 14 liefert wie die Vorversionen sehr gute Ergebnisse in der OCR-Erkennung, sofern die Vorlagenqualität stimmt. Die Layout-Analyse arbeitet so akkurat, dass selbst mehrspaltig gesetzte Texte kor-

rekt erkannt werden. Einschränkungen gibt es bei der Erkennung von Frakturschriften, aufwendigen Tabellen-Layouts oder komplizierten Formeln. Dank der hinzugekommenen PDF-Funktionen und der Möglichkeit zum Dokumentenvergleich mausert sich das Programm zum Allround-Talent fürs Büro. Anders als mit Acrobat, Nuance Power PDF oder Foxit Phantom lassen sich keine PDF-Formulare erstellen.

Wer das Programm zunächst ausprobieren möchte, kann eine 30-Tage-Testversion herunterladen. Ärgern könnten sich diejenigen, denen die Funktionen des Abbyy PDF-Transformer+ ausreichen. Sie bekommen nur noch das Paket inklusive OCR und zahlen dafür einen höheren Preis. Wer ohnehin beide Produkte nutzt, spart nun. Neben der Corporate-Version gibt es eine Standard-Version ohne Möglichkeit zum Textvergleich. In der Enterprise-Version lassen sich mit dem Hot Folder Tool Dokumentenstapel automatisch verarbeiten. (apoi@ct.de/atr@ct.de) **ct**

30-Tage-Testversion: ct.de/ybeu

Abbyy FineReader 14

OCR und PDF-Tool	
Hersteller	Abbyy, www.abbyy.com
Systemanforderungen	Windows ab 7, Windows Server ab 2008 R2
Preise	Standard 200 € (Upgrade 130 €), Corporate 300 € (Upgrade 200 €), Enterprise 500 € (Upgrade 330 €)



Ordentlich Filme schauen

Filmbibliothek ideal für Kodi vorbereiten

Das Media Center Kodi ist auf den meisten Plattformen mit wenigen Klicks installiert. Bis Ordnung in einer großen Filmsammlung herrscht, vergehen jedoch mitunter mehrere Stunden. Tipps und Tools zum schnellen Sortieren.

Von Dennis Schirmacher

Kodi v17 ist da: Eine gute Gelegenheit, Ordnung in die eigene Filmsammlung zu bringen. Das Media Center kann nicht nur unzählige hochauflösende Videoformate abspielen, sondern Inhalte auch optisch ansprechend präsentieren, inklusive Metadaten wie Inhaltsangabe, Cover und Trailer. Damit das klappt, muss man Kodi lediglich mitteilen, wo lokal oder auf Netzwerkspeichern Filme zu finden sind.

Die Informationen bezieht Kodi automatisch online vom Community-Projekt The Movie Database (TMDB). Dabei orientiert sich der TMDB-Scraper

standardmäßig am Namen der Filmdatei. Alternativ kann man in Kodi auch die Option „Filme liegen in getrennten Ordnern, die dem Filmtitel entsprechen“ aktivieren. So oder so ist entscheidend, dass die Bezeichnung korrekt ist. Stimmt diese nicht, wird der Titel mit falschen Infos verknüpft oder taucht erst gar nicht in der Bibliothek auf. Angst vor Lücken des Online-Archivs braucht man nicht zu haben: Die TMDB-Datenbank ist riesig und wir haben alle 200 Titel unserer Test-Bibliothek gefunden.

Filmname reicht nicht

Probleme tauchen auf, wenn es in der Datenbank mehrere Titel mit demselben Namen gibt. Nehmen wir als Beispiel für eine fehlerhafte Verknüpfung den Film „Auf der Flucht“, in dem Harrison Ford den flüchtenden Dr. Richard Kimble mimt. Selbst wenn die Datei den korrekten Titel trägt, verknüpft der TMDB-Scraper die Datei mit den Infos zum Original „Auf der Flucht“ aus dem Jahr 1963; das ist schlicht und ergreifend der erste Tref-

fer für diesen Titel in der TMDB-Datenbank. Der Schlüssel für eine korrekte Verknüpfung ist neben dem korrekten Filmenamen das Jahr, in dem der Titel in die Kinos kam. Erst wenn der Dateiname „Auf der Flucht 1993“ lautet, klappt es mit der richtigen Darstellung in der Bibliothek.

Damit man nach dem Importieren von großen Sammlungen nicht den Überblick verliert, bei welchem Film es zu Fehlern gekommen ist, hilft ein Blick in die von Kodi standardmäßig protokollierten Ereignisse unter „Einstellungen/System“. Dort kann man die den Fehler verursachende Datei ausfindig machen und Gegenmaßnahmen ergreifen. Achtung: Kodi schreibt nach jedem Neustart eine neue Log-Datei. Führt man anschließend keinen erneuten Bibliotheks-Scan durch, tauchen im Log keine falsch indexierten Titel auf. Wer der Übersicht halber gerne eine ausgedruckte Liste in der Hand halten möchte, muss lediglich das Video-Add-on „Missing Movies“ aus dem Kodi-Repository installieren. Nach dem Start

kann man über den Punkt Export eine Text-Datei mit den fehlerhaften Einträgen erzeugen, abspeichern und ausdrucken.

Helferlein

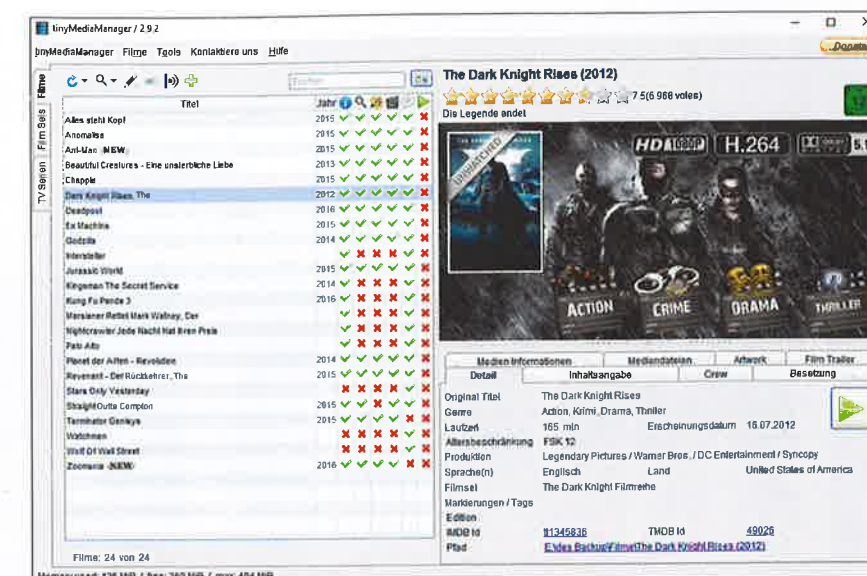
Das Umbenennen klappt am bequemsten über den Datei-Explorer am Computer mit einer Tastatur, kann aber auch direkt aus Kodi über den integrierten Dateimanager geschehen. Bei großen Filmsammlungen artet das jedoch in Fleißarbeit aus. Doch keine Panik: Um diese Arbeit so effizient wie möglich zu erledigen, gibt es diverse kostenlose Open-Source-Tools, die Dateien automatisiert umbenennen.

In die engere Auswahl haben wir MediaElch und tinyMediaManager (siehe c't-Link) genommen, weil es diese Tools jeweils für Linux, macOS und Windows gibt. Beide Werkzeuge beziehen die Infos unter anderem aus der TMDb. Mit tinyMediaManager konnten wir bereits nach kurzer Einarbeitungszeit unsere Filmsammlung optimal für das Einlesen durch Kodi vorbereiten. Beachten sollte man aber, dass tinyMediaManager eine Java-Laufzeitumgebung (JRE) ab Version 1.7 benötigt.

Komfortabel verwalten

Beim ersten Start von tinyMediaManager legt man den Speicherort der Filme fest. Nach einem Klick auf den Aktualisierungsknopf im Hauptfenster des Tools erschienen in unserem Fall alle Titel zur weiteren Bearbeitung in einer Liste, auch die nicht korrekt benannten. Bevor es losgeht, empfiehlt es sich, in den Einstellungen des Tools unter dem Punkt Filme/Allgemein den Haken bei „Automatisch umbenennen“ zu setzen.

Nun braucht man nur noch die unter Kodi nicht korrekt erkannten Filme in der Liste zu markieren und nach einem Rechtsklick die Option „Suchen & Scrapen aller ausgewählten Filme“ zu starten. In den anschließend auftauchenden Fenstern muss man lediglich das jeweils korrekte TMDb-Suchergebnis auswählen und mit „Ok“ bestätigen. Im Anschluss funkt tinyMediaManager das API von TMDb an, lädt automatisch verschiedene Bilder (Poster, Logo et cetera) herunter und packt die Metadaten in eine *.nfo-Datei. Auf Wunsch kann man beides in den Einstellungen unterbinden. Wer



Mit dem kostenlosen tinyMediaManager kann man seine Filmsammlung mit wenig Aufwand optimal für den Import in das Media Center Kodi vorbereiten.

möchte, kann mit dem Tool alle Infos auch manuell bearbeiten.

Unter dem Reiter „Film Dateien umbenennen“ kann man feintunen: Standardmäßig hängt das Tool noch Infos zum Video- und Audiocodec an einen Dateinamen. Wer das nicht will, löscht in dem Reiter einfach die Parameter \$V \$A aus der Zeile hinter „Dateiname“. Praktisch: tinyMediaManager benennt nicht nur Dateien automatisch um, sondern erzeugt auch einen neuen Ordner nach dem von Kodi verlangten Namensschema und kopiert den Titel inklusive *.nfo-Datei und Bildern dorthin – der alte Ordner wird automatisch gelöscht.

Korrektes Ergebnis

Am Ende spuckt tinyMediaManager dem Kodi-Namensschema (Filmtitel plus Jahr) entsprechende Pfade aus. Damit klappt das Einlesen in die Filmbibliothek des Media Centers fehlerfrei. Das geht übrigens auch offline, da Kodi direkt auf die Metadaten in den von tinyMediaManager angelegten *.nfo-Dateien und die heruntergeladenen Bilder zugreift. Dafür muss man in Kodi beim Import unter „Bitte Informationsquelle auswählen“ die Option „Local information only“ anklicken.

Um einen falsch verknüpften Bibliothekseintrag für eine einzelne nachträglich korrekt umbenannte Datei zu aktualisieren, muss man nicht einen Komplett-Scan durchführen. Ein Weg ist zum Beispiel, in Kodi den Punkt „Informationen“ eines fehlerhaften Eintrags in der Biblio-

thek aufzurufen und dort auf Aktualisieren zu klicken. Findet der TMDb-Scraper einen Titel gar nicht, etwa aufgrund eines Schreibfehlers, und taucht dieser nicht in der Bibliothek auf, fügt man die Filmdatei mit korrigiertem Namen entweder manuell per Rechtsklick direkt aus dem Kodi-Dateimanager zur Bibliothek hinzu oder führt einen Komplett-Scan durch.

Da geht noch mehr

Wer schon eine korrekt benannte, aber noch nicht mit Metadaten und Bildmaterial versehene Filmsammlung auf einem Speichermedium liegen hat, kann tinyMediaManager auch komplett automatisiert laufen lassen. Das gelingt über die Option „Alle ungescrapten Filme suchen & scrapen – nimm beste Treffer“. Der Vorgang gerät jedoch schnell ins Stolpern: Bereits ein einziger Buchstabendreher in einem Titel führte in unserem Versuch zu keinem Ergebnis und wir mussten Hand anlegen.

Zusätzlich kann man mit dem tinyMediaManager diverse Untertitel herunterladen. Bereits vorhandene Untertitel benennt das Tool automatisch der Filmdatei entsprechend um – das klappt auch, wenn die Untertitel in einem Unterordner des Filmordners liegen. Dank MediaInfo lassen sich darüber hinaus technische Infos zu Audio- und Videoformaten von Videodateien einsehen. (des@ct.de) **ct**

Tools zum Verwalten von Filmarchiven:
ct.de/yt5n