

Die Ebook-Werft

Ebook-Selbstbau leicht gemacht

Ein Tutorial von MWB
Erstellt im Oktober / November 2023



<https://ebook-werft.de/>

*Unvollständige Ausgabe
(V: 0.7 / Stand 28.10.2023)*

*„Zwei Dinge sind zu unserer Arbeit nötig:
Unermüdliche Ausdauer und die Bereitschaft,
etwas, in das man viel Zeit und Arbeit gesteckt hat,
wieder wegzuwerfen..“*

Albert Einstein

Willkommen!

Hallo lieber Leser,

schön, dass Du hergefunden (oder Dir sogar das Tutorial heruntergeladen) hast!

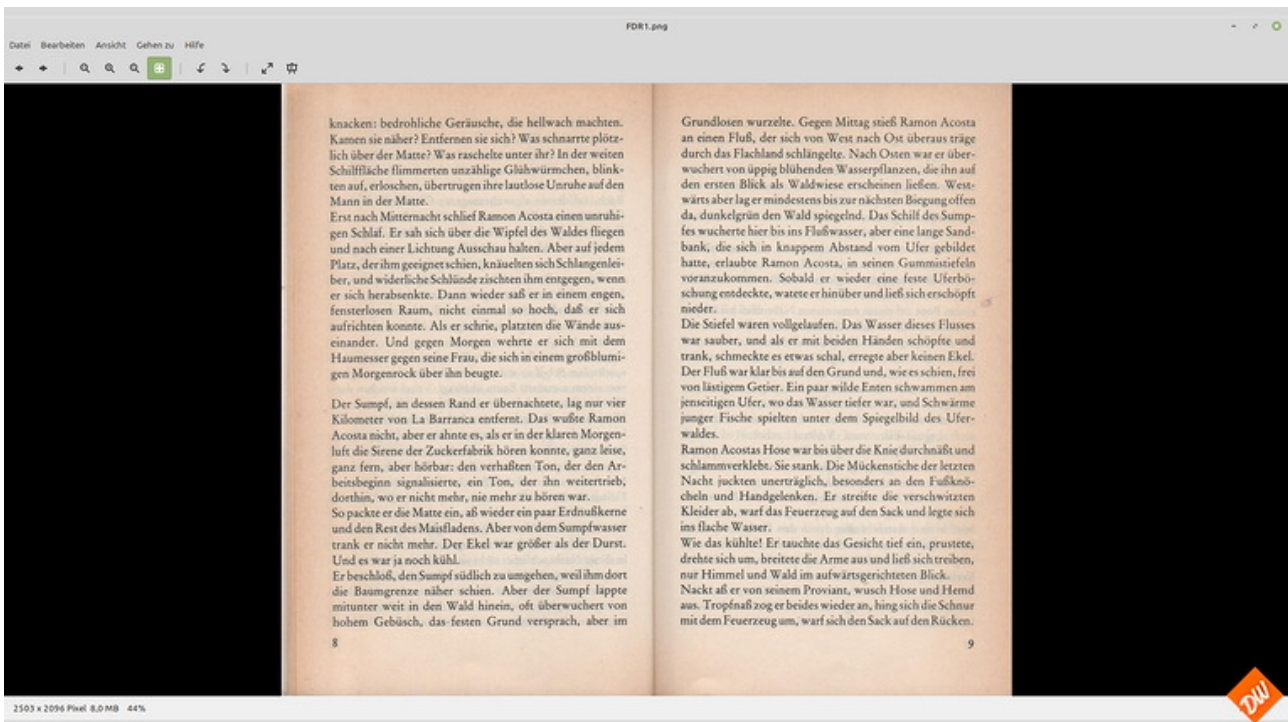
Du wirst vermutlich gute Gründe haben, warum Dich interessiert, wie man selbst am eigenen Rechners Ebooks herstellen kann?

Weil Du zum Beispiel einen eigenen oder gemeinfreien Text als Ebook weitergeben oder gar veröffentlichen willst – oder weil Du es, wie ich vor einigen Jahren, zunehmend unbequemer findest, Deine alten, leider nie in elektronischer Form erschienen Bücher auf Papier zu lesen, noch dazu bei schlechtem Licht und mit Brillengläsern, die dicker und dicker werden?

Wie es begann

Denn so war es damals bei mir, als ich anfang, mich mit diesem Thema zu beschäftigen: Viele alte, nie als Ebook erschienene, zerfledderte und teils schon stark vergilbte Taschenbücher, die ich gerne nochmal gelesen hätte – und Augen, die damit immer mehr Schwierigkeiten hatten, bis ich beinahe völlig die Lust am Lesen verloren hätte.

Deswegen habe ich irgendwann beschlossen, diese Bücher in stundenlangen Sitzungen Seite für Seite einzuscannen, um sie (zuerst in Form einzelner Bild-Dateien, kurz darauf im nächsten Schritt dann schon als PDF des kompletten Buches) wenigstens noch auf meinem Tablet lesen zu können:



Damit ging Lesen dann zwar schon wieder deutlich besser als das hantieren mit den Buchvorlagen, aber als wirklich gut erwies sich diese Notlösung trotzdem noch nicht. Weil ich dabei auch ständig mit einem prinzip-bedingten Nachteil von PDF-Dateien konfrontiert war, die ja im Grunde auch nicht mehr als ein Stapel von Bildern sind, die man in einem Paket zusammengepackt hat:

Man kann halt Bilder nur im Ganzen zoomen muss sehr viel scrollen, wenn die gewählte Ansicht nicht komplett auf den Bildschirm passt. Erst recht ein Problem, wenn man sie auf einem Ebook-Reader mit seinem – verglichen zum Tablet – sehr kleinen Bildschirmformat lesen will.

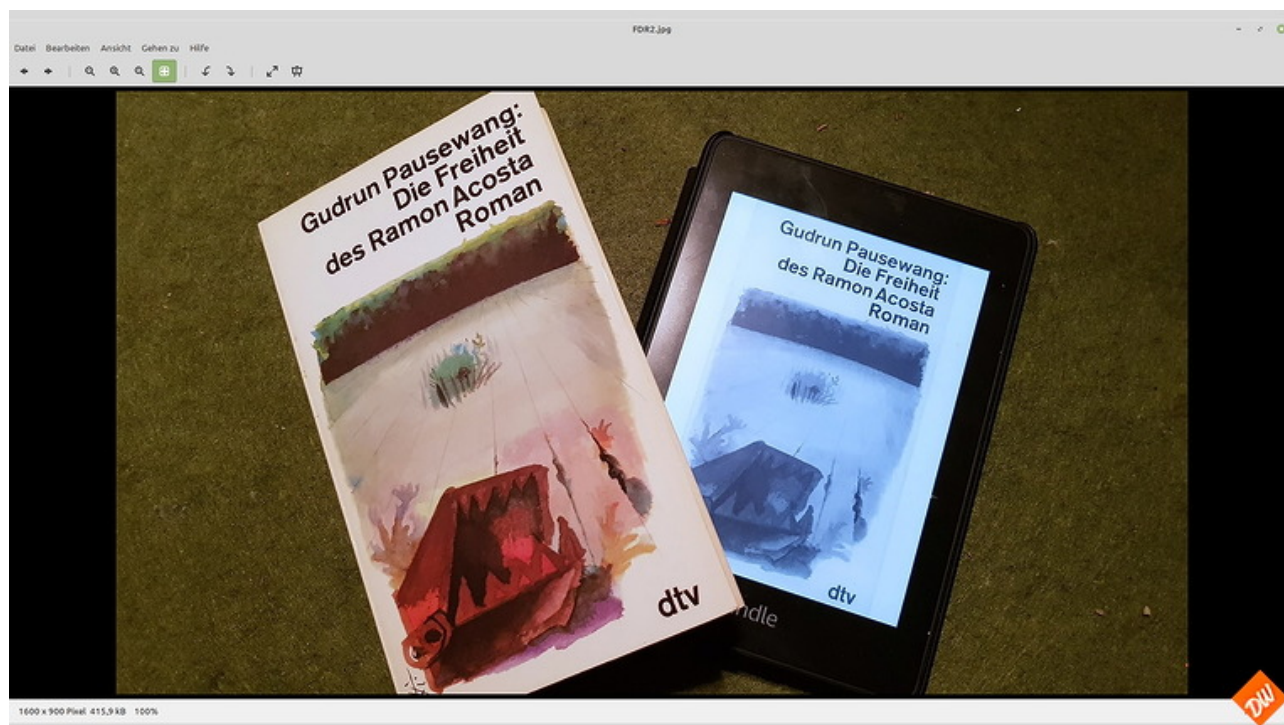
Was mich zu der Frage führte, ob es nicht eventuell doch einen Weg gäbe, meine Scans in ein E-Reader-taugliches Format mit Fliesstext zu bringen, um noch ein wenig angenehmer lesen zu können? Und in der Folge zu ersten Experimenten mit dem führte, was ich an Werkzeugen ohnehin zur Verfügung hatte, mich orientierend an Anleitungen, die ich mir mühselig überall im Netz zusammengesucht hatte...

Immerhin musste ich dafür erst mal nichts investieren:

Zeit genug hatte ich als Teilzeitrentner, einen Computer hatte ich, einen halbwegs brauchbaren Scanner samt Texterkennungs-Programm auch und ein Export ins Epub-Format war auch aus meinen Office-Programm heraus möglich. Warum es also nicht mal versuchen?

Also habe ich mich ziemlich spontan und ohne grosse Vorüberlegung ans Werk gemacht und gleich – mich völlig überschätzend – mit einem Fünfundert-Seiten-Wälzer begonnen, den ich zuvor als Ganzes in eine fette PDF-Datei gepackt hatte.

Keine gute Idee, wie ich schnell feststellen musste, weil mein vorhandenes Equipment damit ziemlich überfordert war und ich auch mit meinen damals noch kaum vorhandenen eigenen Fähigkeiten dabei immer wieder an meine Grenzen geriet. Aber das war meinem ausgeprägten Basteltrieb folgend erst recht ein Anlass, mich mit weiteren Versuchen an das heranzutasten, was möglich war. Mit kleineren Dateigrößen (!), mit viel ausprobieren nach dem Prinzip „Versuch und Irrtum“ – und letztendlich auch zum Erfolg führend mit dem ersten Buch, was lesbar auf meinem Ebook-Reader landete:



Zugegeben, das war bei allem Stolz auf diesen ersten Erfolg von der Gestaltung her noch kein Meisterwerk, wie ich bei der anschliessenden Lektüre feststellen musste, aber es war zumindest lesbar mit allen Vorteilen, die ein E-Reader bietet, allem voran natürlich mit der guten Skalierbarkeit der Schrift und seiner guten Beleuchtung, die Lesen im Bett oder an anderen dunkeln

Orten erst zu einer angenehmen Sache machen.

Und damit war es auch eine Grundlage, mich weiter mit dem Thema zu beschäftigen, zumal da ja noch ein ganzer Haufen an Büchern lag, an deren Papierversion ich lesend gescheitert war – auch dieses Fünfhundert-Seiten-Ungetüm, was den eigentlichen Anstoss zu meinen ersten Versuchen gegeben hatte und das ich gerne in perfekter Form lesen würde....

... und was daraus wurde.

Das gab denn auch zusammen mit meinem Hang zur Perfektion ein paar Bücher später den Anstoss, meinen inzwischen grundsätzlich funktionierenden Workflow noch weiter zu perfektionieren – und auch das noch auszutüfteln, was mir an Fähigkeiten scheinbar noch fehlte:

Vor allem Feinheiten in Textverarbeitung – bei Fehlerkorrektur und Formatierung, wie man Inhaltsverzeichnisse gestaltet, Links in Ebooks setzt, Bilder und Silbentrennung einbinden kann und was darüber hinaus noch beim Export in Reader-taugliche Formate nutzbar sein könnte.

Also war wieder einmal viel Lesen angesagt, wieder viel Zeit im Netz mit Recherche in den Tiefen von Websites und Foren zu verbringen, auf denen es bruchstückhaft Informationen dazu gab. Dabei fiel mir damals schon auf (und es ärgert mich auch heute noch), dass sich wohl im deutschsprachigen Raum noch niemand die Mühe gemacht hat, mal alles an einer Stelle zusammenzufassen, was es zu diesem Thema an Wissenswertem gibt – obwohl es anscheinend eine ganze Menge Leute gibt, die sich damit wenigstens in Teilaspekten schon beschäftigt haben und auch entsprechende Anleitungen anbieten.

Meine Idee zu dieser Seite

Anleitungen, die ich gerne angenommen, gebündelt und genutzt habe bei den inzwischen etlichen dutzend Ebooks, die in der Folge noch an meinem Rechner entstanden sind (auch besagter Fünfhundert-Seiten-Wälzer ist schon lange so bearbeitet und lesend konsumiert worden) – auch wenn dabei die Suche nach Informationen gelegentlich sehr zeitraubend war, oft genug auf Irrwege führte und es dabei manchmal schwer fiel, Spreu vom Weizen zu trennen.

Aber immerhin kam dabei im Lauf der Zeit eine stattliche Linksammlung heraus, mit der ich arbeiten kann, nachdem ich weiß, wo und wie ich suchen muss – und dazu schon vor längerer Zeit die Überlegung, diese auch irgendwie teilen zu wollen.

Oder, noch besser: eine eigene Webseite zu dem Thema zu erstellen, in die auch meine eigenen Erfahrungen einfließen könnten. Zumal ich offensichtlich auch nicht der einzige bin, der sich damit beschäftigt.

Struktur muss sein

Blieb aber noch die Frage, wie ich am Besten an das Thema herangehen könnte, das sich bei genauer Betrachtung doch als recht vielschichtig erweist und zumindest in Teilen auch nicht als sonderlich interessant für Menschen erweisen würde, denen es „nur“ um eigene Texte geht, wohingegen mein Fokus ja eher auf dem Umarbeiten / für mich wieder lesbar machen „fremder Texte“ (aus alten Büchern) liegt, rein zum privaten Gebrauch und nicht unbedingt zur Veröffentlichung bestimmt?

Umfassend und übersichtlich sollte es ja trotzdem sein, schon wegen der Möglichkeit weniger interessante Teile zu überspringen...

Was es hier zu lesen geben wird

Deshalb scheint es mir nach einem ersten Fehlversuch auch am Besten, mich bei der Gestaltung an meinem eigenen Workflow zu orientieren, beginnend mit allem, was dazu nötig ist (also Hard- und

Software) weiterführend mit Kapiteln zum Scannen, mit der Texterkennung und Textbearbeitung und schlussendlich auch mit dem Export der Bücher bis auf den eigenen Ebook-Reader, jeweils aufgeteilt in eine kurze Vorstellung der nötigen Software und praktische Tipps zu deren Verwendung.

Und natürlich auch (in einem abschliessenden Kapitel) mit ein paar Worten zur rechtlichen Situation, insbesondere zum Urheberrecht im Umgang mit fremden Texten.

- - - - -

BTW.:

Parallel zur online gewählten Form als feste, einzeln aufrufbare Seiten sind die bisher veröffentlichten Abschnitte dieser Anleitung auch in der Reihenfolge ihrer Entstehung in Form eines Blogs mit Kommentarfunktion erschienen, um Neufassungen und Veränderungen dokumentieren zu können und eine Dialogmöglichkeit zu bieten

Dies erscheint mir sinnvoller als eine Kommentarmöglichkeit unter den statischen Seiten, auf denen teilweise mehrere Abschnitte zusammengefasst sind. Lob, konstruktive Kritik und Rückfragen (auch zu dieser Form des Tutorials) sind dort natürlich gerne willkommen

Was es hier einstweilen nicht geben wird

Womit ich mich bisher noch nie beschäftigt habe ist das Thema einer Veröffentlichung eigener Texte, zumal ich bisher auch nie die Intention hatte, mich über meine Blogs hinaus selbst schreibend zu betätigen.

Insofern hab ich momentan auch nicht geplant, dieses Thema im Rahmen dieser Seite aufzugreifen, obwohl es meine Anleitung nach ihrer Fertigstellung auch als Download geben soll – zumindest als PDF und möglicherweise auch portiert auf Ebook-Formate.

Und dann könnte es eventuell doch sein, dass es noch ein Zusatzkapitel gibt und das Thema Veröffentlichung relevant wird. Schaumermal....

BTW.:

Auf der Website wird es nach ihrer Fertigstellung auch eine Newsletter-Funktion geben, damit Du ggf. benachrichtigt wirst, wenn es Überarbeitungen, Erweiterungen oder auch Neuauflagen dieses Tutorials gibt

Zu guter Letzt

Bevor es richtig losgeht noch ein kleiner, aber notwendiger Disclaimer:

Alle Inhalte dieser Seite geben nur das wieder, was meinem Wissensstand und meinen eigenen Erfahrungen zum Zeitpunkt der Erstellung dieser Seiten im Oktober/November 2023 entspricht. Was bei mir funktioniert, muss deshalb bei Dir, lieber Leser, nicht unbedingt genauso funktionieren, auch wenn ich denke, dass ich die meisten Eventualitäten dabei berücksichtigt habe. Support auf Deine Fragen kann ich deshalb auch nur im Rahmen dessen bieten, was meine Erfahrungen und meine technischen Möglichkeiten zulassen. Garantien auf Erfolg kann ich nicht übernehmen.

Und bedenke bitte:

Was Du mit den Inhalten dieser Anleitung machst, tust Du immer auf eigene Verantwortung. Du haftest selbst für die Schäden, die daraus entstehen können.

Auch im Zusammenhang mit rechtlichen Problemen, wenn Du Dich nicht an die Regeln der Urheberechtes hältst und nur auf meine laienhaften Angaben verlässt.

- - - - -

Die Bilder auf dieser Seite, insbesondere die Screenshots fremder Texte dienen lediglich Demonstrationszwecken und wurden nicht in der Absicht erstellt, fremde Rechte zu verletzen. Sollte es damit Probleme geben, bitte ich darum, mich über das Kontaktformular unter dem Impressum anzusprechen, damit ich sie ggf. austauschen oder entfernen kann.

Ende der Vorrede und noch eine Schlussbemerkung:

Wenn Du Dich nun an das Experiment wagen willst, meinem Beispiel zu folgen, dann bedenke bitte auch folgendes:

Was Du vorhast, wird neben Hard- und Software auch eine Menge Zeit, Geduld und Fleiss von Dir fordern und kann auch sehr frustrierend werden, wenn sich der Erfolg nicht gleich einstellen will.

Fange also erst mal mit etwas kleinem an, bevor Du Dich an ein grösseres Projekt wagst und sei dabei auch nicht zu streng mit Dir selbst.

Denn auch aus Misserfolgen lernt man auch etwas..... und Meister fallen bekanntlich nicht vom Himmel.

Und nun lass uns das Spiel beginnen, zunächst mal mit etwas Theorie und Grundwissen.

Rechner, Monitor & Co

Hier geht es nun erst mal um das, was Du auf jeden Fall brauchst und wahrscheinlich ohnehin besitzt:

Einen Rechner (oder einen Laptop), einen Monitor und (unbedingt empfehlenswert) eine Mouse dazu als gut handhabares Zeigegerät

Und falls Du gedruckte Vorlagen digitalisieren willst: natürlich auch einen Scanner.

Aber darum wird es erst im nächsten Abschnitt gehen.

Der Rechner

Prinzipiell eignet sich erst mal jeder halbwegs aktuelle Rechner und auch jedes nicht allzu alte Laptop, um Texte für ein Ebook zu bearbeiten und ins passende Format zu bringen. Jedenfalls, solange beide gewisse Mindestvoraussetzungen erfüllen:

Willst Du nur eigene Texte bearbeiten, sind die Anforderung zumindest nicht höher, als sie für ein x-beliebiges Office-Programm und eine ganz einfache Bildbearbeitungen wie Paint nötig wären.

Wenn Du allerdings auf der Grundlage von Scans und Texterkennung arbeiten willst, dann sollte es schon etwas mehr sein. Denn dabei würdest Du an einem alten Rechner mit einem schwächlichen Celeron-Prozessor und 4GB RAM vermutlich verzweifeln, weil Dein Rechner damit schnell an seine Grenzen kommen dürfte, da insbesondere die Texterkennung als Grafikanwendung sehr Ressourcenhungrig ist.

Da sollten es dann schon ein Quadcore-Prozessor der Mittleren Leistungsklasse und wenigstens 8GB RAM sein. Tendenziell auch gerne mehr, wenn Du nicht allzuviel Zeit in irgendwelchen Warteschlangen verplempern willst.

Dein Betriebssystem spielt dabei eher keine Rolle, denn grundsätzlich ist hier egal, ob Du nun lieber Linux, Windows oder MacOS nutzt, solange der Rechner stabil läuft. Einschränkungen gibt es hier nur bei der Frage der Verfügbarkeit von Treibern und Software für Deinen Scanner, bei der Windowsnutzer etwas im Vorteil sind, was das Angebot angeht – und Linux-Nutzer meist das Nachsehen haben, wenn sie nicht viel Zeit in die Konfiguration Ihrer Hardware stecken wollen. Doch dazu später mehr, wenn es um diese Geräte geht.

Wichtig auch: ausreichend freier Festplattenplatz, um temporär notwendige Dateien zwischenspeichern zu können, selbst wenn das fertige Produkt meist nur weniger als ein Megabyte gross ist. Rechne dafür ruhig grosszügig, denn bei dicken Wälzern mit einigen wenigen Bildern hatte ich auch schon mal ein Gigabyte an Daten auf der Festplatte, die zum Glück später wieder gelöscht werden konnten

BTW.:

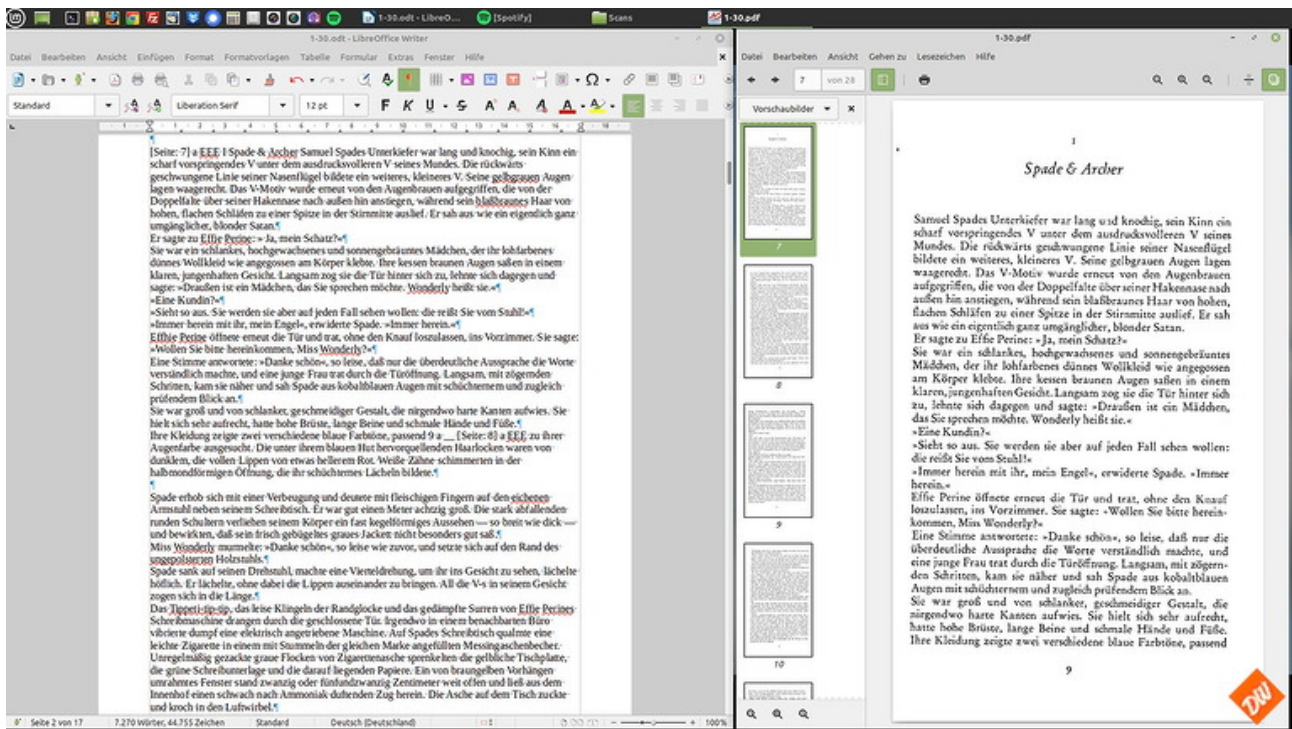
Ich selbst nutze übrigens einen schon sieben Jahre alten ehemaligen Office-Rechner mit einem auf 4×2,9 GHZ getacketen intel-I5-Prozessor, 16G RAM und aktuellem Linux Mint als Betriebssystem – sowie ein hardwaremässig sehr ähnlich ausgestattetes Laptop, auf dem ebenfalls Linux Mint und daneben Windows 11 Professional installiert sind.

Beide Systeme sind mehr als ausreichend für komfortables und zügiges Arbeiten.

Der Monitor

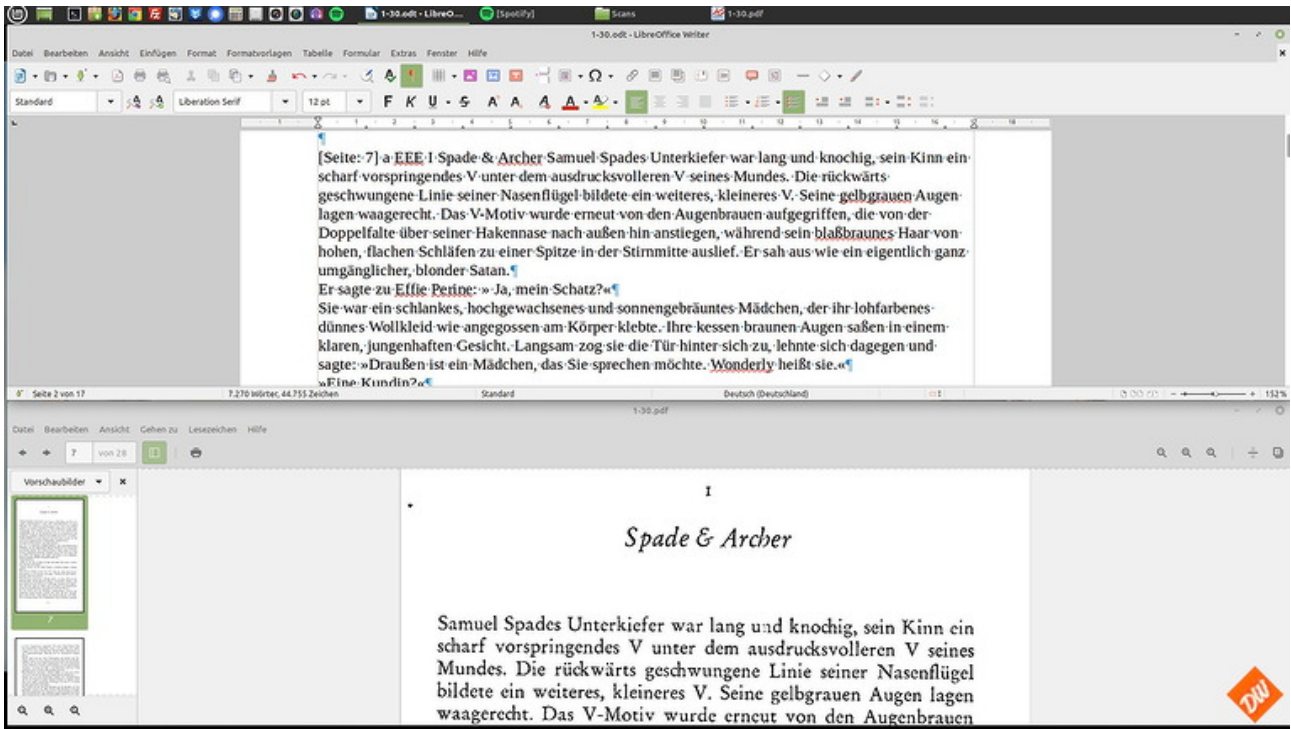
Auch hier geht erst einmal alles, was Du ohnehin schon hast.

Sowohl mit einem Laptop-Bildschirm als auch mit einem separaten Monitor kann man gut leben, sofern er nicht zu winzig ist und ausreichend Arbeitsfläche in der Breite bietet. 1600 Pixel, besser sogar 1920 Pixel sollten es also mindestens schon sein, und auch nicht gerade ein 14-Zöller-Laptopmonitor, wenn Du komfortabel arbeiten willst, ohne Dir dabei die Augen zu verderben. Solche kleinen Bildschirme reichen eventuell, wenn Du damit nur in einem Office-Programm unterwegs bist, aber kaum, wenn Du die Textverarbeitung und die gescannte Vorlage nebeneinander darauf sehen möchtest – wie hier bei der Fehlerkorrektur oder später bei der Formatierung.



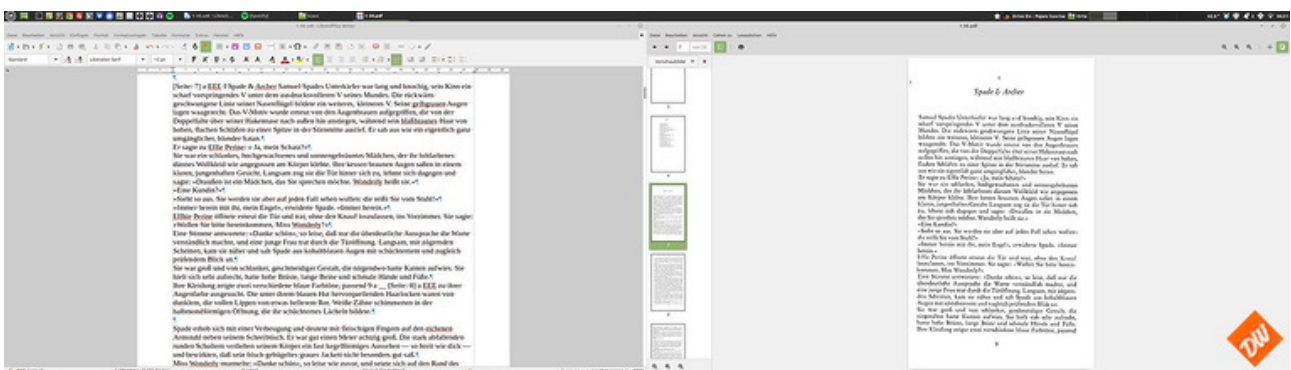
Wobei zwar Du zwar zur Not die beiden Programme auch untereinander anordnen könntest, um diesen Nachteil etwas auszugleichen, das aber mit weniger Übersichtlichkeit erkaufst und deutlich mehr scrollen musst, wenn Du aus Platzmangel nicht gleich die gedruckte Vorlage neben die Tastatur legen willst.

(auch das habe ich schon ausprobiert und hatte als Brillenträger wenig Freude daran)



Deshalb mache ich diese Arbeiten inzwischen auch am liebsten am meinem Arbeitsrechner mit seinem Doppelmonitor-Setting aus zwei nebeneinander montieren 27-Zöllern – und mit der Möglichkeit, die beiden Ansichten in PDF-Reader und Textverarbeitung unabhängig voneinander zoomen und scrollen zu können (oder auch mal die Playlist zu wechseln in dem ich nur ein Fenster minimiere. Musik bei der Arbeit muss sein), ohne an Übersichtlichkeit in einer von beiden zu verlieren.

Hier ist mehr tatsächlich mehr, wenn man gut sehen will – und es spricht auch bei der abschliessenden Formatierung einiges dafür, eine ganze PDF-Seite in der Übersicht zu haben, wenn man sein Ebook ähnlich gestaltet haben will wie das Original.



Aber gut, das ist natürlich eine Frage persönlicher Vorlieben und auch ausreichender Platzverhältnisse auf Deinem Schreibtisch – und sicher auch kein Muss für das, was wir vorhaben.

Die Maus

Aus meiner Sicht – neben einer guten Tastatur – unabdingbar für präzises Markieren von Abschnitten, Worten oder auch nur einzelnen Buchstaben – mag sie für viele versierte Trackpad-

Nutzer ein Greuel sein: Die Mouse, die ich auf meinem Schreibtisch hin und her schubse. Sicher ebenfalls eine Frage der Gewohnheit und persönlicher Vorlieben, völlig egal ob mit oder ohne Schwanz. Hauptsache, man kommt gut damit zurecht und kann schnell und einfach das markieren, was man gerade bearbeiten will.

- - - -

Womit die grundlegenden Hauptkomponenten nun abgearbeitet wären und wir in den nächsten Abschnitten noch einen genaueren Blick auf das werfen sollten, was sicher nicht für die Menschen relevant ist, die Ebooks vor allem aus eigenen Texten erstellen möchten: Scanner und Texterkennung alles, was damit zusammenhängt.

Spätestens ab dem Thema „Textverarbeitung“ dürfte es aber auch für diese Leser wieder interessanter werden

Scanner (I): Übersicht und Gemeinsamkeiten

(Wer vor allem Ebooks aus eigenen oder „gemeinfreien“ Texten erstellen will, kann dieses Kapitel und die weiteren Seiten zum Thema „Scanner und Texterkennung“ ohne Weiteres überspringen)

Nachdem wir im letzten Kapitel die Frage nach Rechner und notwendiger, weil arbeitserleichternder Peripherie geklärt haben, sind nun die Geräte dran, die Du nur dann brauchst, wenn Du auch Ebooks aus gedruckten Texten erstellen willst:

Scanner, also im weitesten Sinne nichts anderes als Digitalkameras, die in der Lage sind, aus einer gedruckten Vorlage ein digitales Bild zu machen, wie Du es als Grundlage für die weitere Bearbeitung benötigst.

Eine erste Übersicht

Prinzipiell kommt dafür auch erst einmal jedes digitale Bildaufnahmegerät in Frage, ob nun Handy aus der Hand oder Digicam auf einem Stativ, Flachbettscanner als Einzelgerät oder in Kombination mit einem Drucker, sogenannte Aufsicht- oder Buchscanner, Duplexscanner mit automatischem Einzug oder sogar hochspezialisierte und extrem teure Faksimile-Scanner, wie manche Bibliotheken sie besitzen.

Allerdings taugt auch nicht jedes dieser Geräte gleich gut für unsere Zwecke, wie sich im Weiteren zeigen wird:

Denn entscheidend ist ja immer, „was hinten dabei rauskommt“, also die Qualität der Bilder, die ausreichen muss, damit im nächsten Arbeitsschritt, bei der *Texterkennung*, die Fehlerquote möglichst niedrig bleibt und wir später, bei der *Textverarbeitung*, möglichst wenig Zeit mit der reinen Fehlerkorrektur verbringen müssen.

Mindestvoraussetzungen der Scans

Insofern sollten die Scan-Ergebnisse auch einige Voraussetzungen erfüllen, die Du mit deinem Gerät erzeugen kannst:

- Eine Auflösung von mindestens 300 Dpi (mehr geht auch, verlängert aber die Zeit für die Texterkennung)
- Ausreichend Kontrast (am Besten reines Schwarz und Weiss, flauere Graustufenbilder und farbige Aufnahmen erhöhen die Fehlerquote)
- Keine Verzerrungen und keine Verschattungen (führen ebenfalls zu einer hohen Fehlerquote)
- gute Bedienbarkeit
- Mindestens USB 2.0 als Verbindung zum Rechner, damit grosse Datenmengen schnell zum Rechner übertragen werden. (Hoch aufgelöste Scanner-Bilder sind immer ziemlich gross – entsprechendes Tempo ist also Pflicht, wenn Du nicht tagelang am Scanner sitzen willst)
- Für Dein Betriebssystem passende Treiber und Hilfsprogramme, die auch in der Lage sind, nicht nur Einzelbilder auszugeben, sondern sie auch gleich optimieren und als mehrseitige PDF-Dateien auswerfen können.

Deswegen können wir gleich mal einige der Bildermacher wieder von unserer Liste streichen, weil sie die notwendigen Kriterien nicht erreichen oder zumindest zusätzliche Arbeitsschritte mit einer Bildbearbeitung nötig sind (Portieren auf Schwarz-Weiss, Entzerren, Zusammenfügen zu

mehrseitigen PDF-Dateien usw...), um aus einzelnen Farb- oder Graustufenbildern eine passende Vorstufe für unsere Texterkennung zu basteln – oder im Falle des Faksimile-Scanners sein hoher Preis und sein Platzbedürfnis dagegen sprechen.

Kurzer Einschub:

Das Format mehrseitiger PDF ist gegenüber Einzelbildern deshalb im Vorteil, weil es eine Menge Arbeitsschritte einspart.

Jeweils 20-40 Buchseiten darin zusammengefasst sind deutlich besser zu handhaben, als wenn man (mal das Beispiel des 500-Seitigen Wälzers aus der Einleitung genommen) Unmengen von Einzelbildern einzeln durch die Texterkennung schleusen müsste, jedesmal den gleichen Arbeitsgang wiederholend.

Andererseits macht es für die weitere Bearbeitung aber auch keinen Sinn, ein komplettes Buch in eine einzige, riesig grosse PDF zu packen, die schon aufgrund ihrer Grösse ewig lange Ladezeiten im Texterkennungsprogramm benötigt und bei der Textbearbeitung viel zu unübersichtlich ist.

Software, Treiber & Kompatibilität

Auch der letzte Punkt der Aufzählung ist wichtig, insbesondere falls Du dir nun erst einen Scanner kaufen willst: Nicht jeder Scanner funktioniert mit jedem Betriebssystem!

Und das gilt sowohl für Windows, wo es für viele ältere Scanner seit der Umstellung auf Windows 11 keine passenden Treiber mehr gibt, als auch für Linux, wo oft die Hilfsprogramme nicht funktionieren, weil es herstellerseitig für dieses Betriebssystem nur in den seltensten Fällen passende Softwareversionen gibt. Lediglich MacOS macht da wohl nicht allzuvielen Probleme, weil sich seit Jahren am grundlegenden Treibermodell kaum etwas geändert hat.

Insofern solltest Du Dich also insbesondere beim Kauf gebrauchter Geräte schon vorab informieren, ob es für Dein Wunschgerät auch passende Software gibt oder wie Du es ggf. mit Work-arounds zum Laufen bekommst. Wobei die erste Adresse immer die Supportseiten des Herstellers sein sollten, aber auch eine Suche mit Google & Co entsprechende Hinweise und Lösungen bringen kann.

BTW:

Keiner meiner drei (!) Scanner funktioniert genuin unter Linux, aber alle liessen sich mit wenig Aufwand und etwas Recherche auf den Seiten [des Sane-Projektes](#) zur Mitarbeit überreden. Dort gibts nämlich [eine Auflistung der unterstützten Modelle](#) und Links zu den passenden Treibern, die von der Linux-Community umgeschrieben und kompiliert wurden. Bezogen auf die Hilfsprogramme sah es allerdings trotzdem eher Mau aus, da half für alle drei nur Ausweichen auf Linux-eigene Lösungen, auf die ich beim Thema Software noch mal näher eingehen werde

Und auch bei Windows kann etwas tricksen hilfreich sein:

Auch zwei meiner drei Scanner werden herstellerseitig unter Windows 11 nicht mehr unterstützt, was passende Treiber angeht. Aber da half in beiden Fällen das Ausweichen auf Treiber für Windows 10 und deren separate manuelle Installation, bevor die zugehörigen Hilfsprogramme installiert werden konnten und nun anstandslos funktionieren. (Übrigens der einzige Grund für mich, doch wieder einen Windows-Rechner zu betreiben, weil die Linux-eigenen Lösungen dagegen deutlich umständlicher zu handhaben sind und weniger Funktionsumfang haben)

Die Hilfsprogramme

Allen Scannern gemeinsam (egal von welchem Hersteller und regelmässig auch auf der Herstellerseite zu finden – leider meist nur für Windows und MacOS. – unter Linux kann man sich beispielsweise mit [Simple Scan](#) behelfen) ist eine Suite mit Dienst- und Hilfsprogrammen, die meist auch selbst schon ein Texterkennungsprogramm enthalten, oft kastrierte Versionen des [Abbyy-Fine-Readers](#) oder von [ReadIris](#) und für unsere Zwecke leider nur eingeschränkt nutzbar, weil beide (im Gegensatz zu ihren kostenpflichtigen Vollversionen) weder die Korrektur der Silbentrennung noch

die Umwandlung in Fliesstext beherrschen, sondern einfach nur stumpf die Zeile um Zeile so aneinander reihen, wie sie erkannt wurden.

Wobei sich zumindest das Problem der Silbentrennung auch mithilfe der Textverarbeitung mit der Funktion "Suchen und Ersetzen" lösen liesse : Wenn man ins "Suchen"-Feld das Trennzeichen mit nachfolgender Leerstelle eingibt und das "Ersetzen"-Feld leer lässt, werden die Trennungen auch entfernt, allerdings erst, wenn vorher auch alle unnötigen Zeilenumbrüche eliminiert sind.

Insofern ist der Weg über die Texterkennung schneller und einfacher

Ansonsten sind diese Suiten immer recht ähnlich aufgebaut und trotz unterschiedlichem Aussehens ohne grosse Unterschiede in der Funktion:

Meist wird bei der Installation gleich ein passendes Symbol auf dem Desktop platziert, über das sich ein Hauptmenue aufrufen lässt, wo man (neben der Möglichkeit, darüber auch andere Geräte des gleichen Herstellers ansprechen zu können) auch die Scanner-Steuerung als Unterpunkt findet.

Klickt man diese an, kann man in weiteren Untermenues die Auflösung einstellen, sowie zwischen Farbscan, Graustufe oder reinen Schwarz-Weiss wählen und ausserdem das Ausgabeformat bestimmen.

Also:

300 DPI – Schwarz-Weiss – mehrseitige PDF

und anschliessend (nach Angabe eines Speicherortes, am besten eines separaten Ordners auf dem Desktop) mit einem Klick auf den entsprechenden Menüpunkt oder Button den Scan auslösen. Screenshots habe ich dazu keine, aber da hilft Dir das Handbuch des Scanners sicher weiter...

Kleiner Tipp:

In der Regel kann man im Scannermenue auch den Kontrast einstellen.

Diese Möglichkeit sollte man unbedingt nutzen, wenn die Scanvorlage farbig oder stark vergilbt ist und sich bei Probescans zeigt, dass die Schrift auf der Seite eher blass wirkt und nicht satt Schwarz erscheint.

Dann kann Nachregulieren an dieser Stelle bei der Texterkennung wahre Wunder wirken

- - - - -

Zum Schluss dieses Kapitels noch eine Anmerkung aufgrund eigener schlechter Erfahrungen

Was Du nicht machen solltest, auch wenn es noch so verlockend sein mag:

Windows-Treiber und Software von einer der zahlreichen, angeblich darauf spezialisierten Seiten herunterladen! Denn damit bekommst Du in der Regel auch nur die gleichen Treiber wie auf den Herstellerseiten und fängst Dir oft noch auch unerwünschtes Beiwerk in Form von penetranter Werbung oder Schlimmerem ein.

- - - - -

Weiter geht es mit der detaillierteren Vorstellung der einzelnen Scanner-Typen.

Scanner (II): Der Flachbettscanner

Sicher am weitesten verbreitet, ob nun als Einzelgerät oder in Kombination mit einem Drucker und teilweise auch als Kopierer und Faxgerät nutzbar, dürfte dieser Scannertyp sein:



Einfach zu bedienen und ein brauchbarer Scanner für viele Zwecke – mit besonderen Stärken beim einscannen flacher Vorlagen wie Bildern oder Briefen:

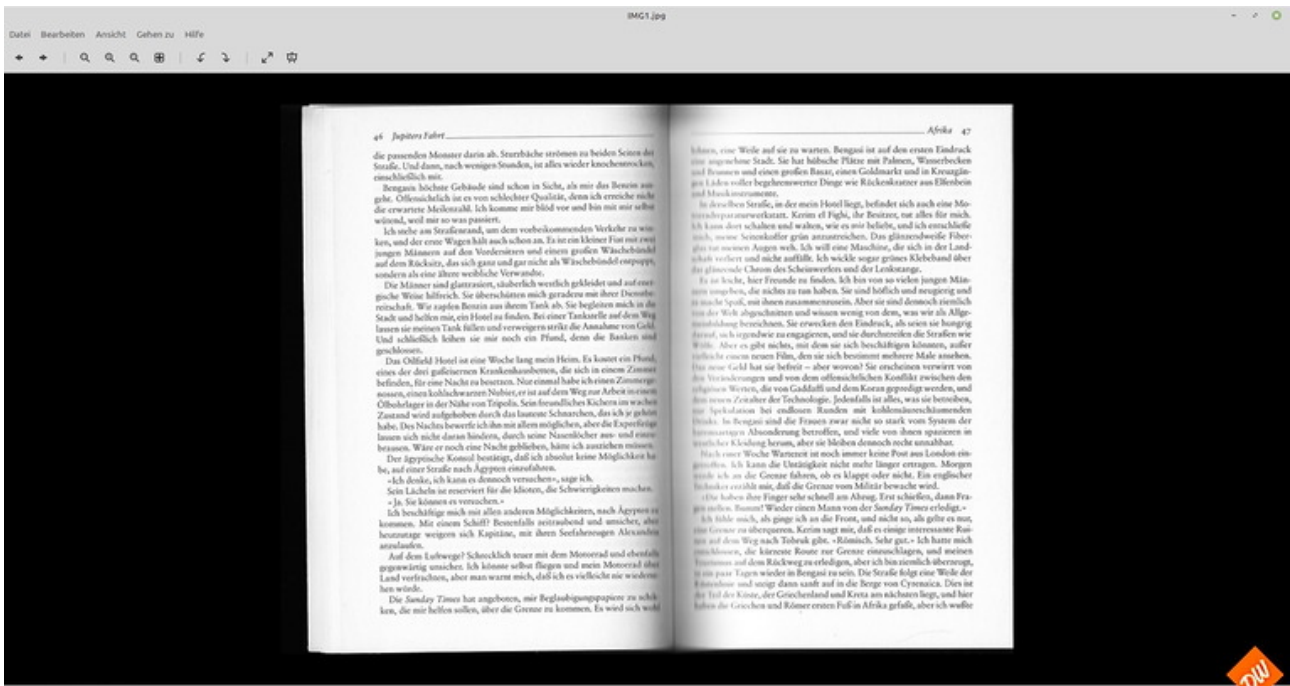
Einfach Deckel auf, Vorlage mit dem Gesicht nach unten auflegen und bsssss...

Und doch ist er beim Scannen von Büchern nicht so ganz ideal, weil zur Prozedur des Scanvorganges dann ja auch immer noch das Umblättern kommt und die Handhabung deutlich umständlicher macht

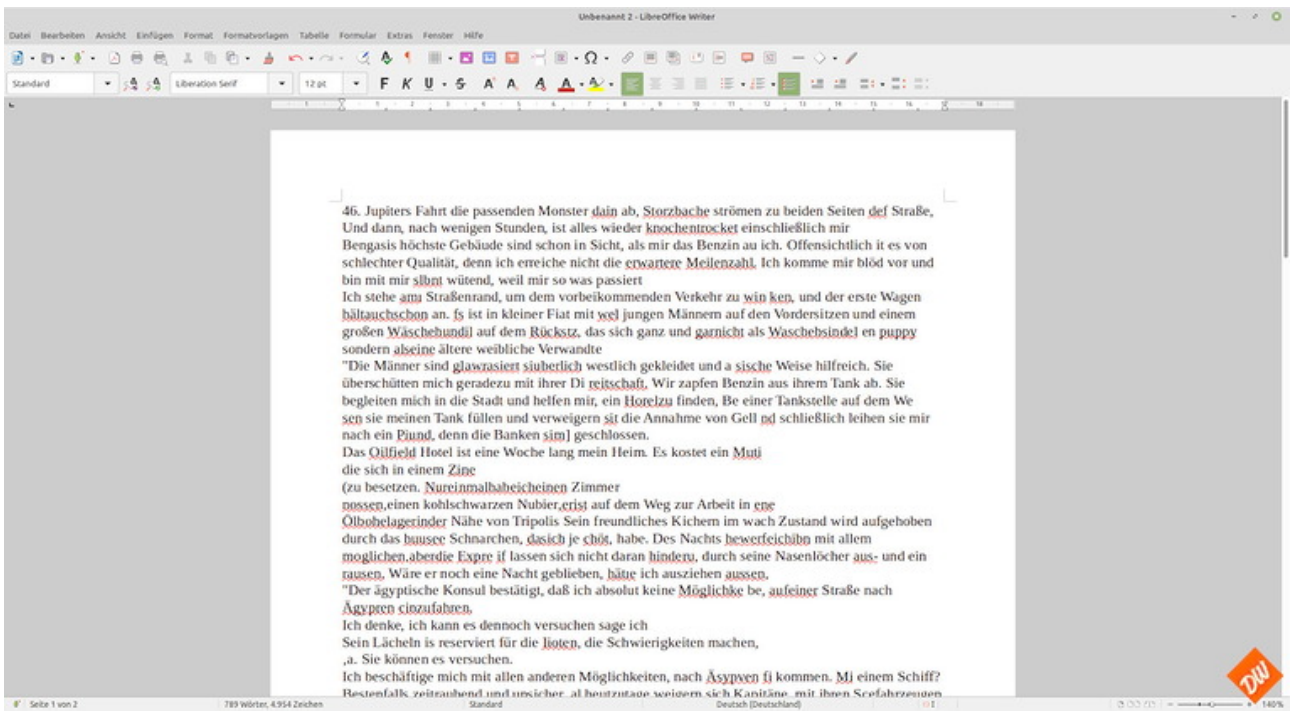
Deckel auf, Vorlage mit dem Gesicht nach unten auflegen, bsssss..., Umblättern, Vorlage mit dem Gesicht nach unten auflegen, bsssss..., Umblättern – usw, usw...

Wozu auch noch zwei prinzip-bedingte Nachteile kommen:

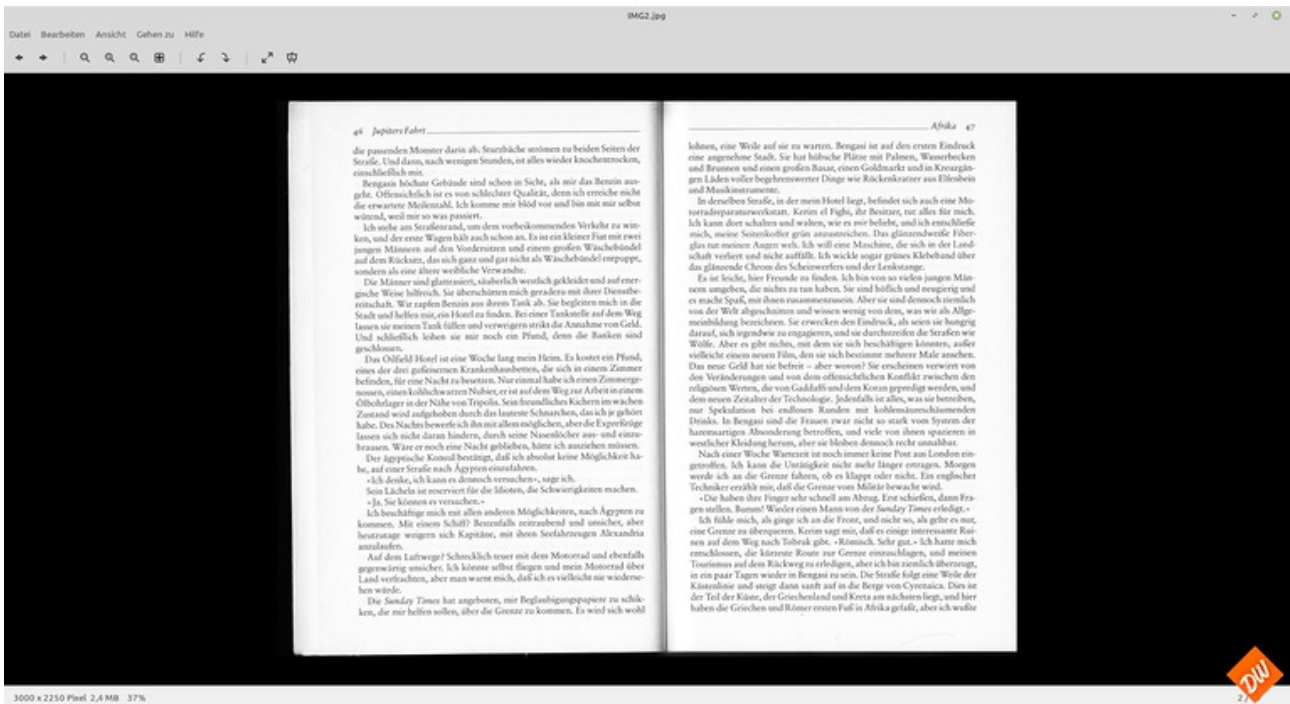
- Zum einen das recht langsame Tempo, weil dieser Scannertyp die Vorlage Zeile für Zeile „liest“ und seine Leseinheit danach erst wieder in die Ausgangstellung zurück fahren muss, ohne dabei etwas sinnvolles zu tun
- Und zum anderen die für unser Zwecke fatale Eigenschaft, dass er nur das wirklich scharf abbilden kann, was sich in direktem Kontakt zur gläsernen Auflagefläche befindet – schlimmstenfalls mit Auswirkungen, die für unser Zwecke nicht zu gebrauchen sind:



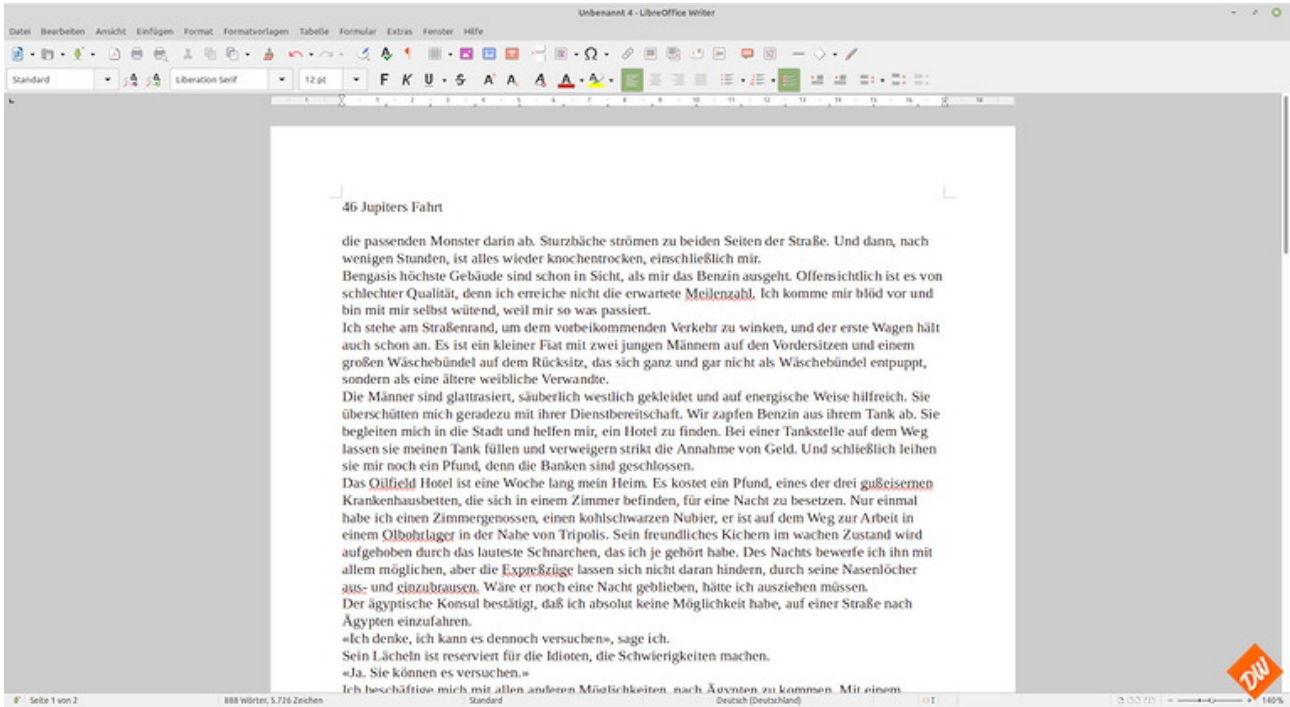
Denn daraus entsteht bei der Texterkennung natürlich nur noch Wortsalat, wie sich später beim Blick auf die Textverarbeitung zeigt, in der es von rot unterstrichenen Fehlern nur so wimmelt, die man mühselig korrigieren müsste:



Doch zum Glück entstehen solche Scanfehler meist nur auf einigen Seiten in der Mitte des Buches und lassen sich auch weitgehend vermeiden, wenn Du genug Druck auf den Buchrücken ausübst, dass die Seiten möglichst komplett auf der Glasscheibe aufliegen. Nicht mit einem Gewicht, sondern am Besten bei offenem Deckel mit den Fingern. Dann sieht das Scan-Ergebnis nämlich deutlich besser aus:



Perfekt für unsere Zwecke, wie ein erneuter Blick in die Textverarbeitung zeigt, wo tatsächlich nur noch Fehler bei Wörtern aufscheinen, die das Wörterbuch noch nicht kennt.



So gesehen kannst Du also ohne weiteres mit diesem Scanner-Typ brauchbare Ergebnisse erzielen, wenn Du mit seinen kleinen Nachteilen leben kannst, die ich vor allem in der umständlichen Scan-Prozedur und in seinem recht langsamen Arbeitstempo sehe- und darin, dass man peinlich darauf achten muss, dass die Scans in allen Bereichen ausreichend Schärfe haben.

Zudem kannst Du Flachbettscanner auch sehr preisgünstig gebraucht erwerben. Auf den einschlägigen Portalen findest Du immer eine reiche Auswahl, auch an Kombidruckern mit defektem Druckwerk, aber noch gut funktionierenden Scannern. Die bekommst Du teilweise schon für den sprichwörtlichen Apfel und das Ei – und gelegentlich sogar ganz und gar geschenkt.

Allerdings solltest Du auch im Schnäppchenfieber nicht vergessen, vorher abzuchecken, ob Dein Wunschgerät auch kompatibel zu Deinem Betriebssystem ist, wie schon im vorhergehenden Kapitel angesprochen.....

Scanner (III) Der Aufsicht- oder Buchscanner

Dieses Gerät mit dem Habitus einer modernistischen Schreibtischlampe hingegen dürfte Dir, lieber Leser, vermutlich völlig unbekannt sein – und auch in meinem Scannerpark ist es ein absoluter Exot:



Schon der Technik wegen die drin steckt – angelehnt an professionelle Faksimile-Scanner, wenn auch deutlich einfacher aufgebaut: Im Grunde eine simple Digitale Kamera (mit einem angepassten Objektiv und parallel dazu eingebauter, regelbarer Beleuchtung), mit der die Vorlage nicht zeilenweise eingelesen, sondern lediglich in hoher Qualität fotografiert wird. Ausgelöst entweder automatisch oder durch ein zugehöriges Fusspedal, das ganz einfach per USB angeschlossen wird: Und so geht es mit bis zu einer Doppelseite pro Sekunde (falls Du selbst so schnell blättern kannst – ich kann es zumindest nicht) ausgesprochen zügig, selbst dicke Wälzer in handliche PDF zu verwandeln:

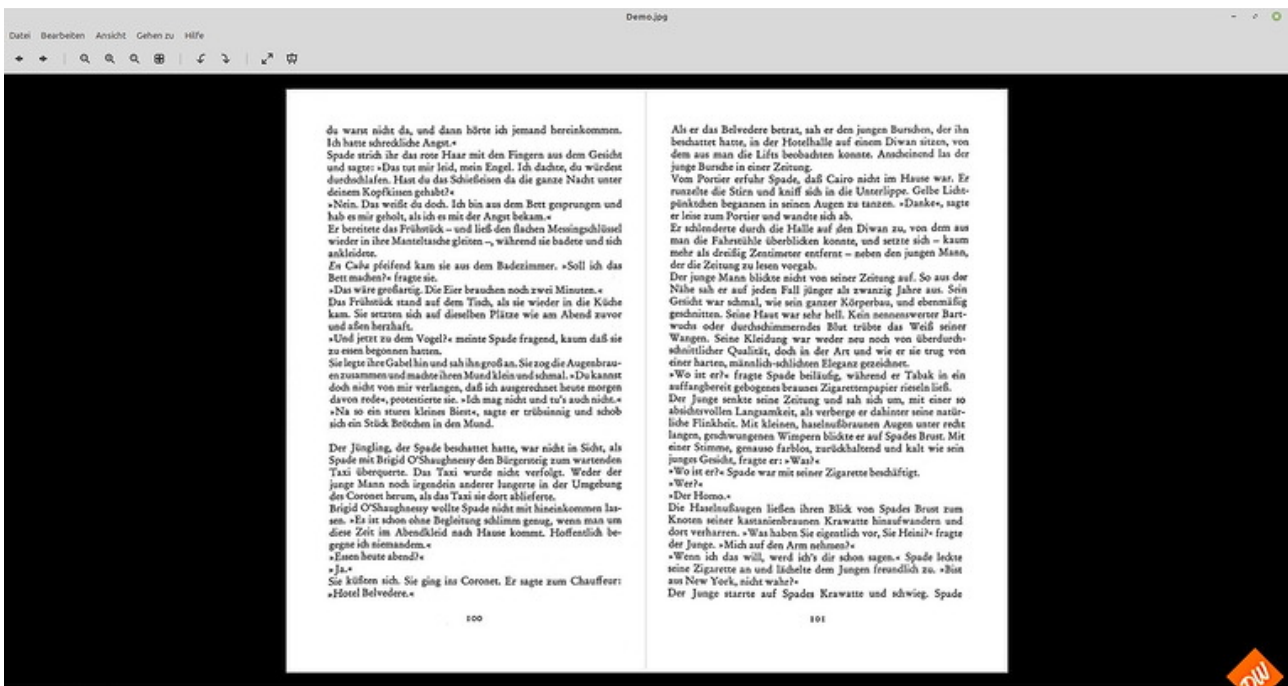
*Einfach Umblättern – aufs Pedal treten – Umblättern – aufs Pedal treten – Umblättern – aufs Pedal treten – usw....
Und anschliessend die Scans direkt als PDF exportieren. Bequemer geht es fast wirklich nicht.*

Wobei der Clou dieses Scanners tatsächlich in seiner (auch hier wieder nur für Windows und MacOS tauglichen Software) steckt, mit der die Bilder im weiteren Verlauf perfekt aufbereitet und in die verschiedensten Formate exportiert werden können (auch als mehrseitige PDF, wie wir sie für unseren weiteren weiteren Workflow brauchen):

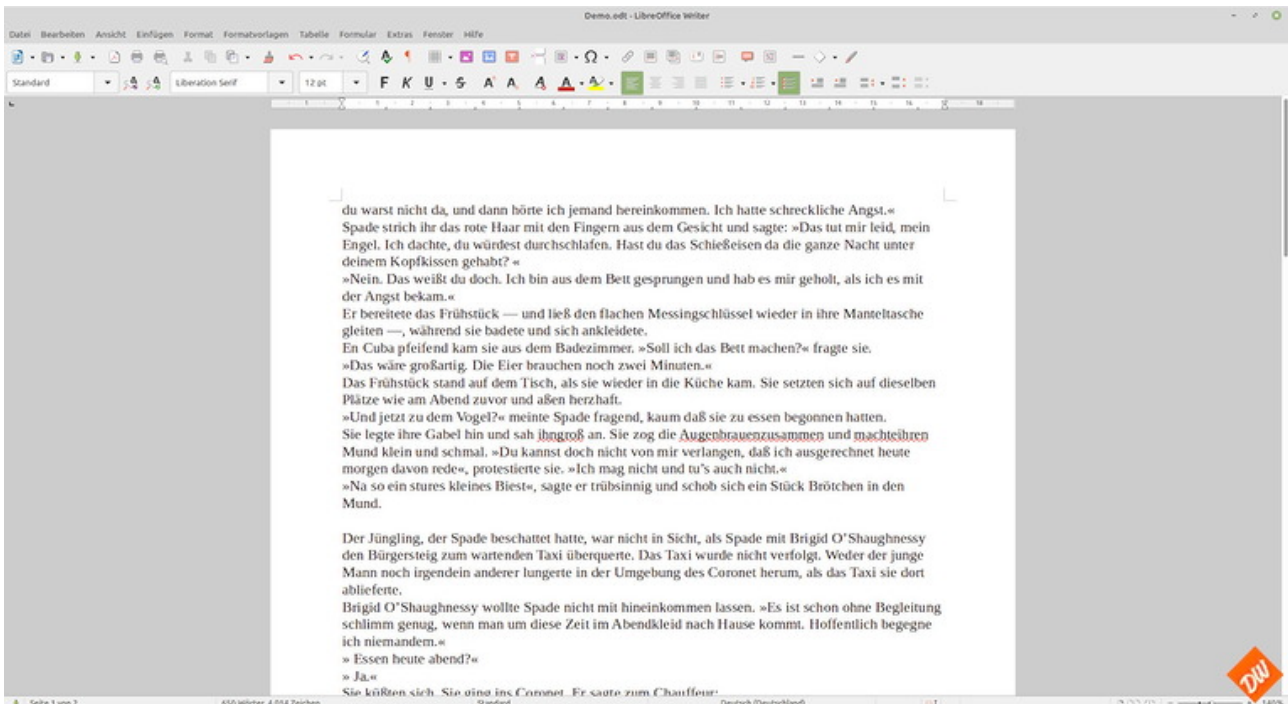
So werden zumindest bei meinem Gerät (einem der mittleren Preisklasse) die Seiten automatisch ausgerichtet, Krümmungen bereinigt und selbst die Finger weitgehend herausgerechnet, die man bei störrischen Buchseiten benötigt, um sie unten zu halten.

Und wenn nicht, schaffen weisse Baumwollhandschuhe wirksame Abhilfe

Entsprechend gut sehen dem entsprechend auch die Ergebnisse aus, wenn sie am Ende der Prozedur abgespeichert werden – hier aus Demonstrationsgründen nachträglich zur Doppelseite zusammen montiert, normalerweise stehen in der Ausgabedatei die einzelnen Seiten untereinander:



Was sich natürlich auch bei der Texterkennung bemerkbar macht, insbesondere, nachdem ich alle Eigennamen schon korrigiert habe, die das Wörterbuch natürlich auch vorher nicht kannte:



Denn hier bleiben tatsächlich als einzige Fehler einige Worte, die in der Vorlage schon so eng zusammen gerückt waren, dass die Texterkennung sie nicht mehr als einzelne Wörter identifizieren und entsprechend voneinander trennen konnte.

Und einen echten Nachteil dieses Scannertyps will ich auch nicht verschweigen:

Für Scans von hochglänzenden Vorlagen oder Fotos taugt er nur bedingt, weil die Beleuchtung direkt neben der Kamera sitzt und sich dabei gerne in der Vorlage spiegelt.

Das wiederum kann ein Flachbettscanner eindeutig besser.

Dennoch ist dieser Scanner für mich schon nahe am Optimum dessen, was sich für unser Hobby noch rechnet, zumal es die billigeren Modelle dieses Typs inzwischen tatsächlich schon zu Neupreisen gibt, die kaum noch über dem eines guten Flachbettscanners liegen.

Ob diese Billigscanner dann aber auch so empfehlenswert sind wie mein gebraucht erworbenes Modell der höheren Preisklasse, kann ich Dir nicht sagen, zumal dabei wohl auch gerne an der mitgelieferten Software gespart wird...

Auf jeden Fall würde ich Dir empfehlen, Dich mit Scannern dieser Art mal näher zu beschäftigen, wenn Du mehr als nur ein oder zwei Bücher scannen willst. Weil aus meiner Sicht die Vorteile überwiegen und er auf Dauer auch viel Zeit bei der Nachbearbeitung einsparen kann

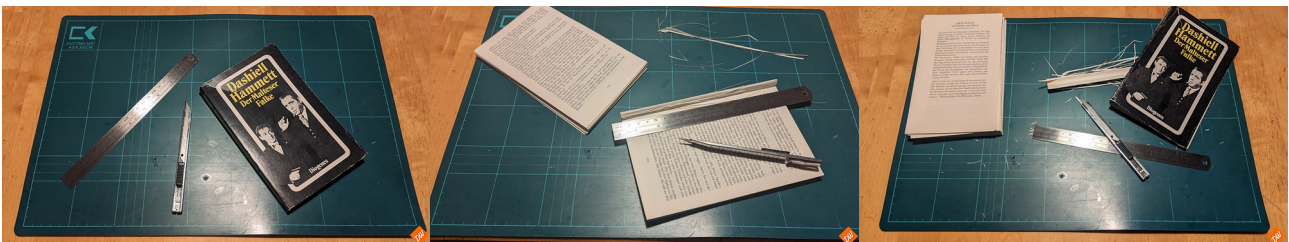
Scanner (IV): Der Duplexscanner mit automatischem Einzug

Dieser Scannertyp, üblicherweise in Büros oder Arztpraxen beheimatet, dürfte mit Abstand (auch gebraucht, weil sehr gesucht) der teuerste der vorgestellten Typen sein und wird sich für Dich wohl nur rechnen, wenn Du entweder vorhast, mit Ebook-Basteleien in Serienproduktion zu gehen, oder ihn wie ich zu einem sehr, sehr günstigen Preis gebraucht erwerben kannst.



Was sich durchaus lohnen könnte, denn er ist tatsächlich in der Lage sehr schnell und sehr ordentlich eine grosse Menge an Vorlagen beidseitig in guter Auflösung einzuscannen (meiner schafft bis zu fünfzig Blatt = 100 Seiten in der Minute!!) und als fertige PDF auszugeben.

Allerdings solltest Du dann auch bereit sein, ein wenig chirurgische Vorarbeit zu leisten und Dein Buch mit scharfem Bastelmesser und Stahllineal vorher zu zerschneiden:

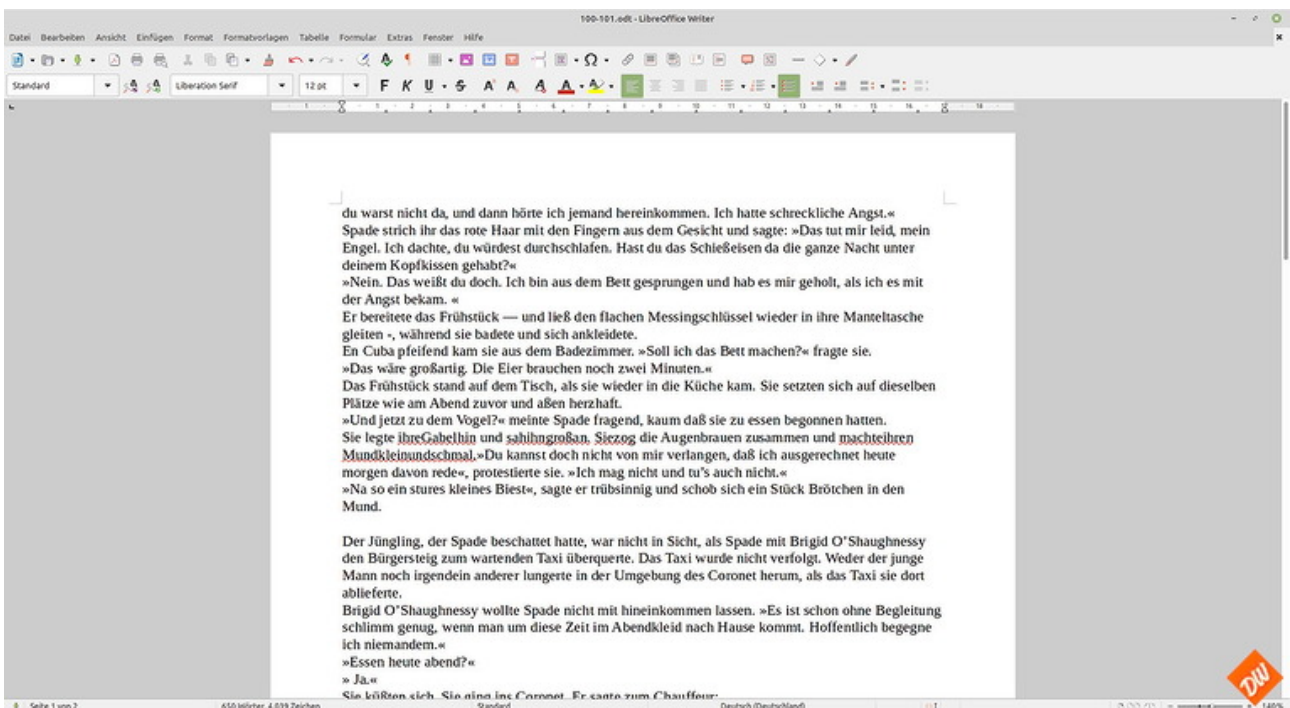
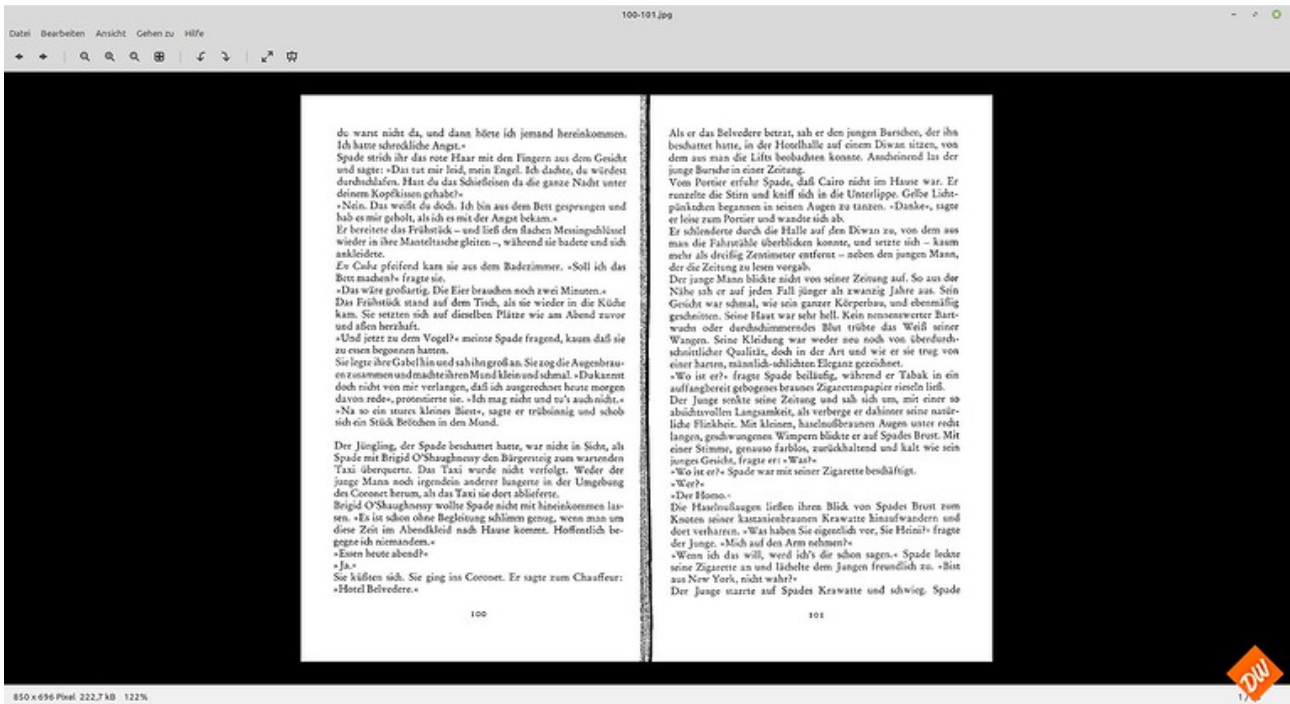


„Am Stück“ kann dieser Scanner aufgrund seiner Bauweise ein Buch oder eine Zeitschrift nämlich nicht verdauen, sondern er braucht tatsächlich die einzelnen Blätter als Futter für seine Arbeit, möglichst so beschnitten, dass die Klebekanten der Leimung *ordentlich* entfernt sind.

Da ich selbst aber hauptsächlich sehr günstig erworbene Letzthand-Taschenbücher in Ebooks verwandele und danach ohnehin ins Altpapier entsorge, weiss ich diese Prozedur als Arbeitserleichterung wirklich zu schätzen und mache mir diese notwendige Mehrarbeit gerne, auch wenn sie manch Anderem als Sakrileg („man zerstört keine Bücher“) oder als anderweitig grosser Nachteil dieses Scannertyps erscheinen mag.

Wobei dieser Scanner tatsächlich alles verarbeiten kann, was entsprechend vorbereitet in den Einzugschacht passt:

Alte schmaddelige und zerlesene Taschenbücher ebenso wie farbige Hochglanzmagazine – mit Ergebnissen, die mindestens ähnlich gut (und meist sogar noch wesentlich besser) sind als beim vorher vorgestellten Bücherscanner:



Und auch bei der Texterkennung erweist er sich als über jeden Zweifel erhaben und hatte mit meinen Testseiten die niedrigste Fehlerquote von allen bis auf (wieder einmal!) die zu dicht zusammenstehenden Worte der Vorlage, mit denen die Texterkennung offenbar auch hier nicht zurecht gekommen ist:

Wobei ich Dir trotz aller Vorzüge dieses Scanner-Typs aber auch noch einen weiteren Nachteil nicht verschweigen will, der zumindest dann relevant werden könnte, wenn Du Dir so ein Gerät gebraucht beschaffen willst:

Tatsächlich gibt es bei diesem Typ zwei oder drei Verschleissteile, die gelegentlich (bei meinem nach 50.000 Scans) gewechselt werden müssen und je nach Fabrikat auch nicht ganz billig sind. Die Andruckrollen für den Papiertransport nutzen sich halt irgendwann ab, sind aber zum Glück kinderleicht zu wechseln und auch für ältere Scanner meist noch problemlos über Amazon zu bekommen...

Frag also vor dem Gebrauchtkauf-Kauf eines solchen Scanners auf jeden Fall nach, ob diese Rollen schon gewechselt wurden und wie viele Scans der Zählerstand in der Scannersoftware aktuell zeigt...

- _ _ _ -

Und noch eines bleibt zu ergänzen, was ich schon mehrfach in diesem Zusammenhang gelesen habe:

Manche Kombi-Drucker haben ja ebenfalls eingebaute Scanner mit automatischem Einzug.

Diese funktionieren trotz komplizierterer (weil um die Ecke führender) Papierführung wohl auch für unser Zwecke, haben aber oft keine echte Duplexfunktion, so dass man schlimmstenfalls den gleichen Papierstapel zweimal hindurchlaufen lassen muss:

Einmal für die Vorderseite und dann nochmal für die Rückseite, um anschliessend die Scans per Software automatisch in die richtige Reihenfolge zu bringen...

Womit sich unter Umständen eine höhere Fehlerquote in der Seitensortierung ergibt und Du zumindest den Zeitvorteil wieder verlierst, den echte Duplexscanner bieten – mal ganz abgesehen von der möglicherweise schlechteren Qualität der Scans, weil die Auflösung dieser Art Scanner dem angestrebten Tempo und dem günstigen Preis geschuldet in diesem Modus manchmal viel zu niedrig ist.

Insofern sind derartige Kombi-Scanner also tatsächlich nur bedingt empfehlenswert, selbst wenn sie wohl meistens ihren Zweck erfüllen.

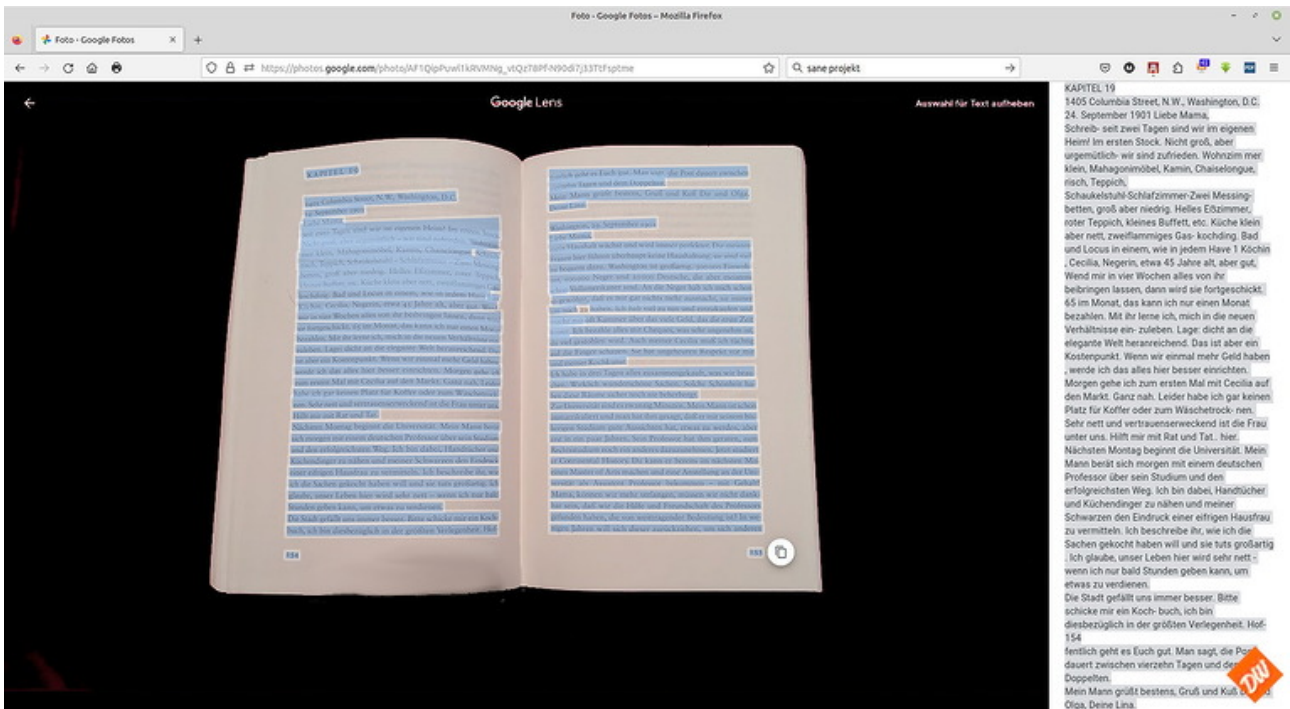
Scanner (V): Die Scanbox

Der Vollständigkeit halbe möchte ich Euch auch noch eine weitere Möglichkeit zum Bücher scannen vorstellen: Diese merkwürdige Konstruktion, hergestellt aus billigen Kunststoffplatten, aber auch in Pappe oder Holz unschlagbar günstig auf allen einschlägigen Onlineplattformen zu finden:

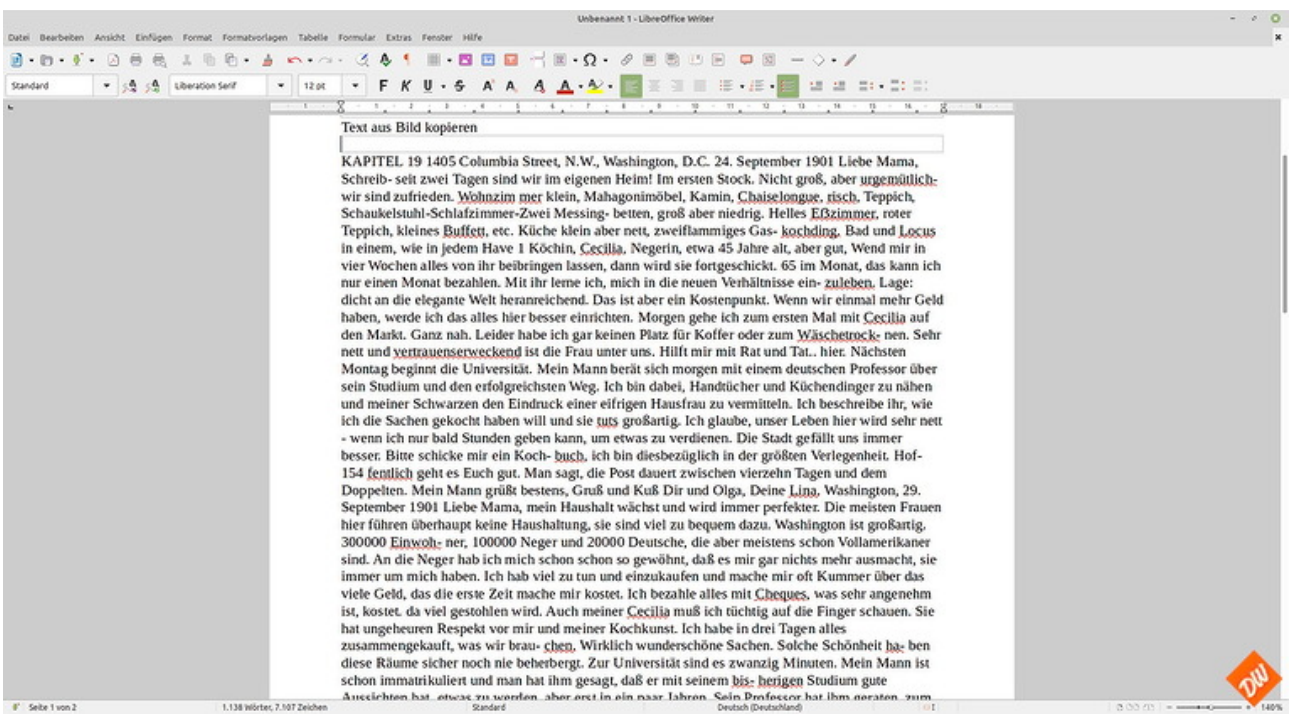


Die Scanbox , auch unter anderen Handelsnahmen erhältlich – und mit ein wenig Eigenleistung auch kinderleicht aus einer alten Bananenkiste aus dem nächsten Supermarkt herzustellen
Ein „Gerät“, was prinzipiell die Funktion eines Aufsicht-Scanners imitiert und dabei das Handy als Kamera nutzt – mit allen Nachteilen, die sich daraus ergeben.

Grund genug für mich, einmal auszuprobieren, was damit denn so möglich ist – und einerseits tatsächlich ein Grund zum staunen, als ich die Texterkennung von [Google Lens](#) nutzte um aus freihand geknipsten Handybildern Textdateien zu machen, die sich direkt in der Textverarbeitung weiter verarbeiten lassen. Denn auf den ersten Blick war die Erkennungsrate tatsächlich ziemlich hoch.



Andererseits aber – und auf den zweiten Blick betrachtet – zeigen sich in der Textverarbeitung neben den üblichen Fehlern bei unbekanntem Vokabeln noch weitere zum Teil gravierende Mängel, die für den weiteren Workflow nichts Gutes verheissen:



Nicht nur, dass Google Lens alle Silbentrennungen ohne Korrektur lässt und gelegentlich Leerstellen in den Text einbaut, wo keine hingehören, sondern auch, dass es den gesamten Inhalt dieser Doppelseiten in einen einzigen Fliesstext verarbeitet und dabei Überschriften und Absätze gnadenlos ignoriert, so das man spätestens bei der abschliessenden Formatierung den gesamten Text durchgehen (und damit wohl auch lesen) muss, um das wieder gerade zu biegen. Doppelseite für

Doppelseite, die man anschliessend auch noch in ein grösseres Dokument zusammenführen muss. Mehr Seiten auf einmal schafft man mit dieser Methode nicht.

Womit diese Lösung sich schlussendlich als ziemlich untauglich erweist und eigentlich nur bestätigt, was ich schon auf einer vorhergehenden Seite über die Tauglichkeit von Handyfotos geschrieben habe. Zusätzlich bestätigt durch eine Gegenprobe mit fest an ein Stativ montierten Handy samt optimaler Ausleuchtung, die trotz etwas besserer Qualität der Bilder ebenfalls bei der Texterkennung sang- und klanglos unterging. Sowohl bei der Nutzung von Google Lens, als auch (mit etwas besserem Ergebnis) mit nachbearbeiteten und optimierten Bildern innerhalb meines üblichen Workflows, die offenbar trotzdem nicht die für die Texterkennung optimale Auflösung von 300 Dpi erreichten.

Wozu dann als weiteres Handicap dieser Handy-Methode auch noch kommt, dass einem in diesem Setting gelegentlich eine dritte Hand fehlt, nämlich dann, wenn die Buchseiten in der Mitte des Buches die Tendenz haben, sich an den Rändern immer wieder hochklappen zu wollen und man seine Finger zum Niederhalten der Seiten benötigt. Dann bleibt im Zweifelsfall nur die Nasenspitze zum Auslösen oder der meist erfolglose Versuch, dasselbe per Sprachbefehl über den Assistenten im Handy anzustossen.

Beides nur sehr unbefriedigende unzuverlässige Lösungen gegenüber dem teureren echten Aufsichtsscanner mit seinem Fusspedal und seiner gut funktionierende Software – und einmal mehr unterstreichend, dass Hilfsmittel wie Handy und Scanbox allenfalls als Notlösung taugen und auf Dauer sicher keine Freude machen werden.

So gesehen sind also selbst die +- 20 Euro noch zu viel, die für derartige Konstruktionen verlangt werden... zumal man mit etwas Glück und Geduld zum gleichen Preis auch einen guten gebrauchten Flachbettscanner bekommen würde, der ausserdem noch einiges an hilfreicher Software mitbringt, auf die man bei dieser Handylösung verzichten müsste.

Jedenfalls lieferte eine weitere Gegenprobe mit meinem alten Flachbettscanner trotz seiner konstruktionsbedingten Macken mit den gleichen Seiten des gleichen Buches deutlich bessere Ergebnisse als das Beste aller Handybilder.

Ein paar Gedanken zum Thema Software

Bevor wir uns dem praktischen Teil dieser Anleitung zuwenden möchte ich Dir noch kurz aufzeigen, welche Software ich abgesehen von [Betriebssystem](#) und [Scanner-Software](#) dafür verwende und auch Dir empfehlen würde.

Wobei ich – von Linux her kommend und partiell auch Windows nutzend – immer Open-Source-Varianten bevorzuge, die (mit Ausnahme des Texterkennungs-Programmes) auf allen Plattformen laufen und möglichst einfach zu installieren und zu bedienen sind.

*Als kleinen Service für Dich habe ich in den einzelnen Abschnitten Download-Möglichkeiten für Windows und MacOS verlinkt – bezogen auf Linux kommt man am leichtesten **an alle hier vorgestellten Programme**, wenn man den Weg über [wiki.ubuntuusers.de](#) geht und dort die Suche bemüht. Dort findet man auch Hilfe zur Installation und Konfiguration der Programme unter Linux*

Die Texterkennung mit einem OCR-Programm

ist nach dem Scannen der wichtigste Schritt, um uns eine gute Basis für alles weitere zu schaffen und gleichzeitig auch der Teil, der als grafiklastiges Programm dabei am meisten Hardware-Ressourcen benötigt. Um so mehr, je umfangreicher die beim Scannen erzeugten Bild- und PDF-Dateien sind.

Also ist es um so wichtiger, dafür ein Programm zu wählen, das selbst möglichst wenig an Ressourcen benötigt und nicht noch unnötigen Ballast mitbringt als viele kommerzielle Programme, deren Vollversionen dennoch geeignet sind, wenn Dein Windows- oder Apple-Rechner ausreichend Reserven dafür hat.

Aber das kam damals für mich als nur-Linux-Benutzer natürlich nicht in Frage, weshalb ich letztendlich bei einer Kombination aus dem von Google gepflegten [Tesseract](#) als OCR-Basis und [gImageReader](#) (nur für Linux und Windows verfügbar) als grafischer Benutzeroberfläche gelandet bin – ähnlich, wie sie bei einigen Online-Angeboten und bei Google Lens zur Texterkennung üblich sind.

Wobei die meisten Online-Lösungen zwar durchaus auch nutzbar sind (damit habe ich anfangs auch herumprobiert) und sogar gute Erkennungsraten haben, aber (ähnlich den kastrierten OCR-Versionen in der Scannersoftware) häufig ebenfalls Beschränkungen in der Funktion aufweisen. Korrektur von Silbentrennungen und Umwandlung in Fliesstext beherrschen nur die wenigsten, oder wenn, dann nur in kostenpflichtigen Abo-Modellen, die unbegrenzten Zugriff erlauben und auch die häufig auf diesen Seiten geltenden Upload-Limits aufheben.

Allerdings wüsste ich auch keinen Grund, auf derartige OCR-Lösungen zurückgreifen zu müssen – es sei denn, mein eigener Rechner käme bei der Texterkennung leistungsmässig an seine Grenzen.

Für Nutzer von MacOS habe ich mangels eigener Erfahrungen an dieser Stelle leider keine Empfehlung ausser, die grossen Suchmaschinen zu befragen.

Die Textverarbeitung

ist das Programm, mit dem Du am meisten Zeit verbringen wirst.

Zumindest bei der hoffentlich nicht allzu notwendigen Fehlerkorrektur und beim sehr zeitaufwändigen Formatieren ist sie unumgänglich und auch sehr hilfreich, wenn Du einen Text

„nur“ als PDF zur Verwendung am Rechner oder auf dem Tablet aufbereiten willst, weil sich hier schon eine direkte Exportmöglichkeit bietet.

Meine Empfehlung dazu ist auf alle Fälle das kostenlose und auf allen Plattformen verwendbare Open-Source-Programm [LibreOffice](#), das im Bereich der Textverarbeitung einen ähnlichen Funktions-Umfang wie Microsoft Word bietet und zusätzlich über Funktionen verfügt (beispielsweise auch einen direkten schnellen Export in EPUB-Format und erweiterte Bearbeitungsmöglichkeiten für PDF-Dateien) die Word nicht kennt.

Falls Du schon Word auf dem Rechner hast, kannst Du das natürlich trotzdem benutzen. Meine Tipps sollten eigentlich auch damit funktionieren

Die Bildbearbeitung

ist ein notwendiges Übel, auch wenn es nicht zur Kernkompetenz von E-Readern gehört, Bilder darzustellen. Aber zumindest ein Titelbild gibt es immer und auch im Text kommen sie gelegentlich vor – und deshalb musst Du Dich zwangsläufig auch mit diesem Thema beschäftigen um sie entsprechend anpassen zu können. Zumindest die Grösse und Schärfe der Bilder solltest Du anpassen können und – weil es der bessern Darstellung dient – auch das Konvertieren von farbigen Bildern in Graustufen.

Immerhin: Als Nutzer von Windows oder MacOs reichen dafür die Bordmittel aus, welche diese Betriebssysteme ohnehin mitbringen – und wenns ein wenig mehr sein darf, gerne auch das ebenfalls für alle Plattformen erhältliche und kostenlose [Gimp](#), wie es in jeder Linux-Distribution ohnehin enthalten ist.

Boliden wie Photoshop sind jedenfalls nicht nötig.

Der Export in E-Reader-taugliche Formate

ist der letzte Schritt zum fertigen Ebook und auch der, wo der letzte Feinschliff stattfindet und alles zusammengeführt wird, was Du in den vorhergehende Schritten erarbeitet hast – Dein korrigierter und formatierter Text und ggf. die Inhalts-Bilder in ihrer angepassten Form.

Auch Titelbild, Inhaltsverzeichnis und Metatexte (Klappentext und Impressum eines Ebooks) werden erst an dieser Stelle erstellt und eingebunden, sowie auch (falls gewünscht) die dynamische Silbentrennung.

Und das alles mit nur einem Werkzeug, mit dem kostenlosen Open-Source-Programm [Calibre](#), für das es leider nur eine englische Website gibt, auch wenn seine Oberfläche perfektes Deutsch beherrscht, die im Übrigen auch noch eine Reihe weiterer, sehr nützlicher Funktionen beinhaltet:

Angefangen bei einer eigenen Bibliothek zur Ebook-Verwaltung (sowohl auf deiner Festplatte als auch auf Deinem E-Reader) bis hin zum komfortablen Ebook-Versand per Mail versammelt sich da alles, was man sich nur wünschen kann, auch wenn ich es hier nicht bis ins letzte Detail beschreiben werde und mich nur auf die Funktionen beschränke, die es für unser Vorhaben braucht.

Weitere Bearbeitung nach dem Export in Ebook-Formate

Auch wenn Calibre schon einiges an Werkzeugen bietet, kann es sinnvoll sein, noch ein weiteres Werkzeug in der Hinterhand zu haben, das allerdings eher für Experten geeignet ist, die schon etwas Erfahrung im Umgang mit HTML und CSS haben:

[Sigil](#) (ebenfalls Open – Source und für alle Plattformen erhältlich.) ist dann hilfreich, wenn man mit den einfachen Formatierungsmöglichkeiten der Textverarbeitung nicht mehr weiterkommt, was insbesondere für Spezialaufgaben wie das Einbinden und die spezielle Formatierung von Bildern zutrifft.

Für „nur Text“-Ebooks mit allenfalls ein oder zwei Bildern wirst Du es also sicherlich nicht brauchen – deshalb erwähne ich es auch nur der Vollständigkeit halber und ohne die Absicht, im weiteren Verlauf dieser Anleitung detaillierter darauf einzugehen. Zumal ich selbst mit diesem Problem bisher nur einmal konfrontiert war und deshalb auch kaum Erfahrung damit habe..

Hilfen dazu gibt es aber einige im Netz – Google wird also auch in Bezug darauf sicherlich Dein Freund, wenn Du Dich an solche Aufgaben heranwagen willst.

Was Du nicht brauchen wirst

sind extra Programme zum Lesen und Bearbeiten von PDF-Dateien, auch wenn insbesondere Windows gerne mal mit diesem Vorschlag um die Ecke kommt. Denn zum Lesen kannst du ganz bequem Deinen Webbrowser nutzen, der mit entsprechendem Plugins (für [Firefox](#) und [Chrome](#) erhältlich) auch das Ebook-Format EPUB darstellen kann (die Kindle-Formate MOBI und AZW3 aber leider nicht) – und zum Bearbeiten reicht Deine Textverarbeitung (egal ob LibreOffice oder Word) allemal aus, solange es nur um das Einfügen oder Löschen einzelner Seiten geht.

Deshalb muss es auch nicht der Adobe-Acrobat-Reader mit seinen teuren Abo-Modellen sein.

Und jetzt kommt der Punkt, wo ich Dir mal gratulieren muss,

weil Du bis hierher durchgehalten hast.

Denn am Ende dieser Seite bist du auch am Ende der grauen Theorie angekommen und kannst anfangen, Dich mit der Praxis zu beschäftigen.

Also: Herzlichen Glückwunsch und auf zu neuen Taten!

Workflow (I): Vorüberlegungen und Vorbereitungen

Nachdem dem Exkurs zur notwendigen Hardware und zur Software – und bevor endlich der praktische Teil dieser Anleitung beginnt – noch ein paar kurze Vorüberlegungen, die ich eigentlich immer vor einem neuen Projekt anstelle.

Schliesslich will ich mir ja auch keine unnötigen Arbeiten machen sondern nur soviel Aufwand treiben wie für den späteren Verwendungszweck unabdingbar ist.

Die „einfache“ PDF-Datei

So muss ich aus einer einfachen Gebrauchsanweisung für ein Haushaltsgerät (soweit die nicht ohnehin schon als PDF downloadbar ist) sicher kein aufwändig zu bearbeitendes Ebook erstellen. Dafür reicht auch einfaches Scannen der gesamten Vorlage in einem Arbeitsgang in niedriger Auflösung (150 DPI reichen für diesen Zweck locker) – ggf. auch noch mit einer schnellen Texterkennung durch die OCR-Funktion der Scannersoftware, um in der PDF später nach Stichworten suchen zu können – und speichern an einem entsprechenden Ort samt Entsorgen der Vorlage ins Altpapier (man muss ja nicht jedes Blatt Papier aufheben), um zum Ziel zu kommen.

Viel mehr Aufwand würde ich auch für Vorlagen nicht treiben, bei denen es im Wesentlichen auf Bilder ankommt oder die ich ohnehin wegen der grösseren Darstellung lieber auf dem Tablet oder am Rechner lesen würde.

Zeitschriften oder Comics beispielsweise, oder Bildbände, deren Original sein Leben als Staubfänger verbringt und allenfalls alle Jubeljahre mal in die Hand genommen wird, Fotoalben, aber auch Fachbücher, die ich nicht gemütlich im Bett eingekuschelt lesen muss – also alle Druckwerke, die zur Verwendung als Ebook nicht taugen oder nicht lohnen – sind mit 150-300 DPI und farbig eingescannt ebenfalls eher Kandidaten für diese einfachste Form der Umarbeitung in ein elektronisches Medium, wobei ich auch hier von Fall zu Fall die OCR-Funktion der Scanner-Software nutzen würde, soweit mir das für eine Stichwortsuche nötig scheint – und darüber hinaus, falls ich auch ein Inhaltsverzeichnis brauche, die Scans später in der LibreOffice noch in Kapitel aufteilen.

Das geht übrigens am Einfachsten, indem man die Vorlage gleich kapitelweise scannt, die Einzelteile später in separate Writer- (oder Draw- je nachdem) -Dokumente einfügt, die man jeweils am Ende mit einem Seitenumbruch versieht und danach zu einem einzigen Dokument zusammenfügt, das man beim Export endgültigen Transport zurück ins PDF-Format auch gleich durch Reduktion der Auflösung noch etwas komprimieren kann.

Darauf gehe ich im Kapitel „Textverarbeitung“ nochmal ausführlicher ein.

Das „richtige“ Ebook

Mehr Aufwand (und damit die im Folgenden beschriebene Herangehensweise) wird also tatsächlich nur für das nötig, was später auch auf E-Readern lesbar sein soll.

Und in etwas reduzierter Form auch dann, wenn, wenn die schnelle PDF-Lösung von der optischen Anmutung her nicht dem entspricht, was wir haben wollen und sich für eine „schöne“ Fassung der Mehraufwand mit Texterkennung, Korrektur und Formatierung lohnt.

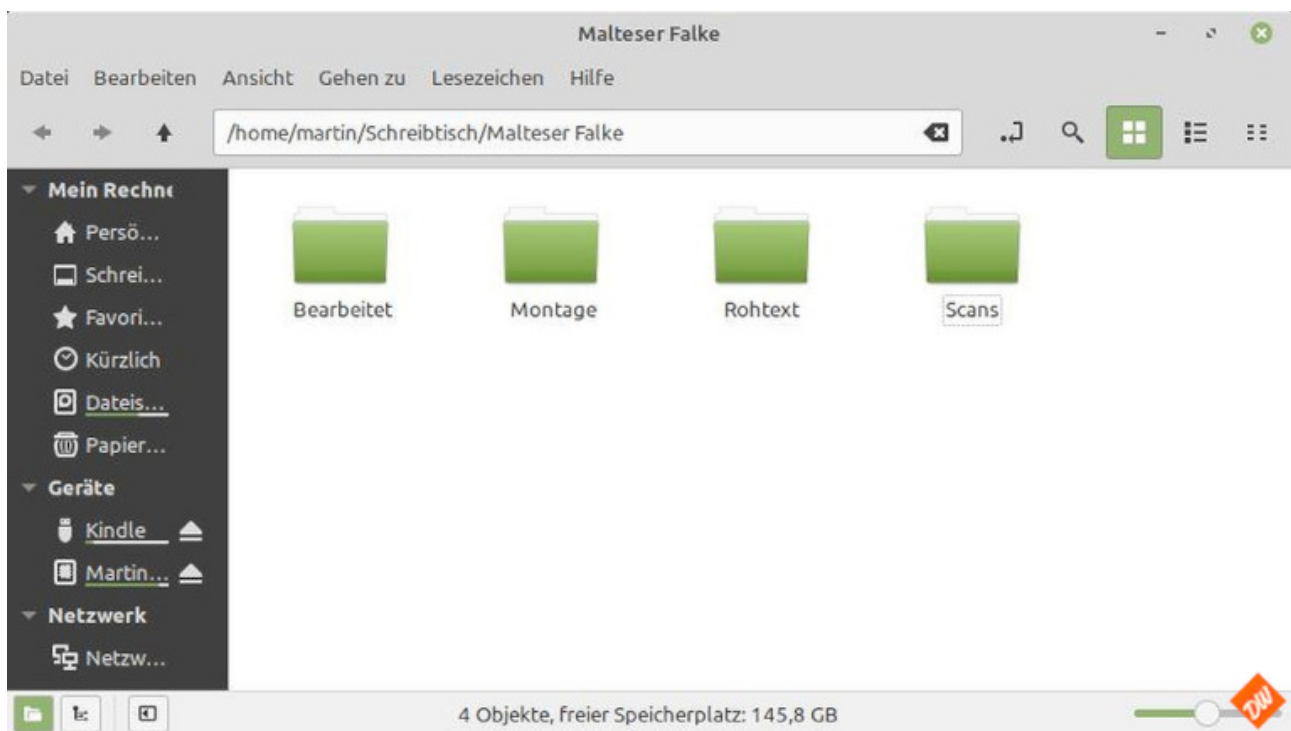
Wobei Bildbände als Vorlage für weiteres Bearbeiten zum Ebook oder für eine bearbeitete Form der PDF so oder so ausscheiden, weil Bilder innerhalb einen Fliesstextes eine Menge Probleme bringen und zudem der Qualität wegen auch vorher noch einzeln bearbeitet und optimiert werden sollten. Wozu ja auch noch die Problematik der eingeschränkten Bild-Darstellungsmöglichkeiten auf E-Readern kommt, wie wir später im Bildbearbeitungs-Kapitel noch sehen werden.

Letzte Vorbereitungen

Nachdem nun auch die Entscheidung über die weitere Vorgehensweise getroffen ist, Kaffee und Snacks bereitstehen (und falls Du Raucher bist: auch genug Tabakwaren für eine länger Sitzung) fehlt eigentlich nur noch eines:

Eine Ordnerstruktur, mit der Du gut arbeiten kannst, denn Du wirst – je nach Umfang der Vorlage – im Verlauf der einzelnen Arbeitsschritte eine ziemliche Menge an zwischengespeicherten Dateien ansammeln, bis am Ende Dein Ebook fertig ist – und dafür auch einiges an Speicherplatz benötigen.

Wobei ich mir meist einen Ordner auf dem Desktop anlege, um mir längeres Suchen zu ersparen – und darin auch gleich die notwendigen Unterordner anlege:



- **„Scans“** (der Name erklärt sich vermutlich von selbst) für alle PDF und Bilddateien die ich für das Ebook brauche, auch für Titelbild und den Klappentext des Buches, den wir später sinnvollerweise in die Metadaten des Ebooks einfügen.
- **„Rohtext“** für die zwischengespeicherten Ergebnisse der Texterkennung, die ich genau wie die Scans immer bis zum Ende meines Workflows aufhebe
- **„Bearbeitet“** für die fertig gestellten Teile aus der Textverarbeitung
- Und **„Montage“** für das, was ich nach und nach daraus zusammenfüge. Auch hier bewahre ich die einzelnen Zwischenschritte auf, bis die finale Fassung erstellt ist.

So hat es sich jedenfalls für mich bewährt, aber natürlich steht Dir frei, Dir auch ganz andere Strukturen anzulegen, wie sie Dir am besten erscheinen.

- - - - -

Bleibt nur noch die Vorbereitung Deiner Scan-Vorlage, mit der dann auch das nächste Kapitel beginnen wird....

.... und der ernst gemeinte Rat, nicht gleich mit dem dicksten Wälzer Deiner Büchersammlung anzufangen, sondern vielleicht erst mal mit einem kleineren und einfach gestaltetem Übungsobjekt zu beginnen.

Workflow (II): Das Scannen

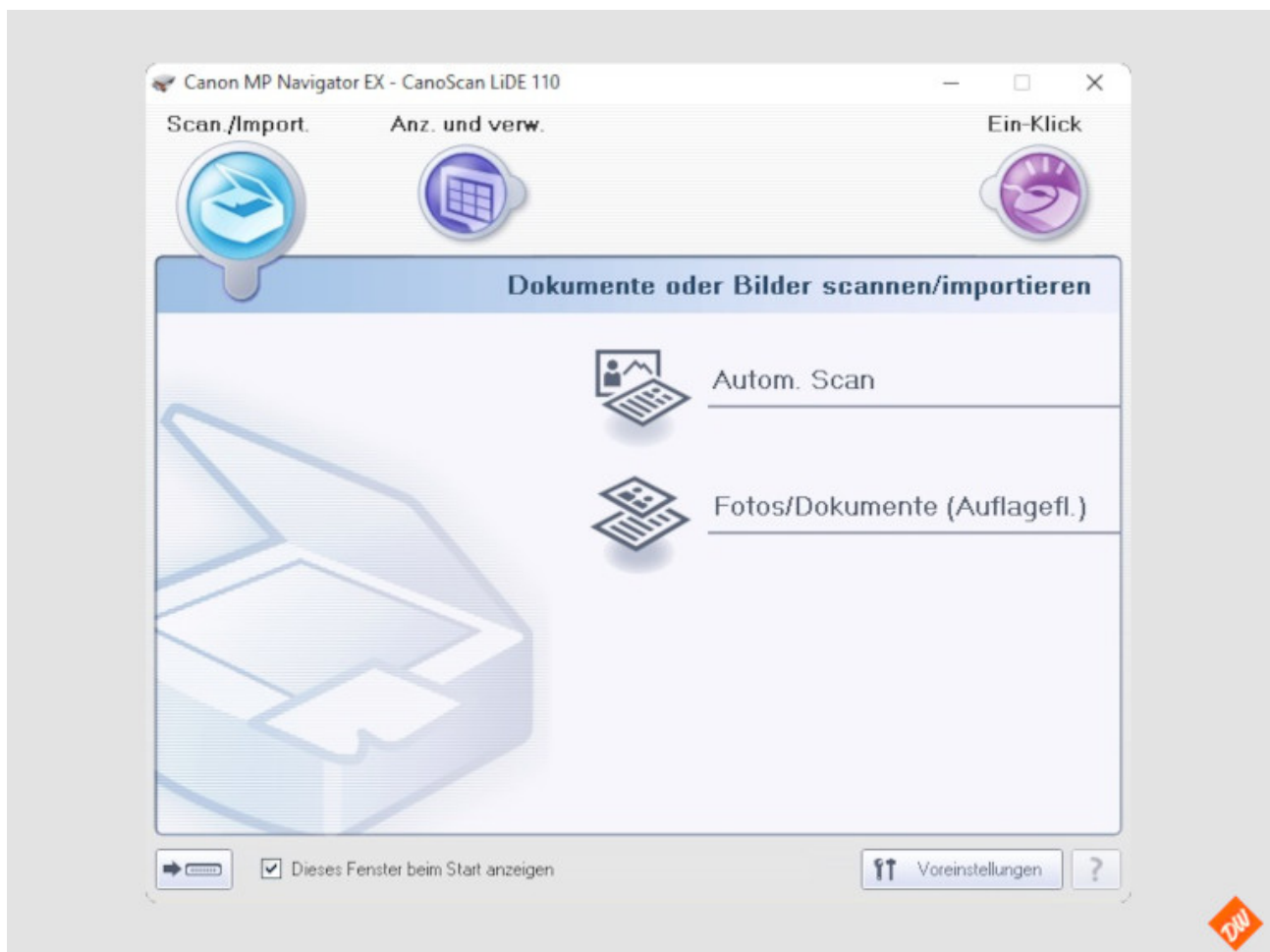
Wichtig vorab

Achte darauf, dass Deine Scanvorlagen und die Auflage Deines Scanners sauber sind. Staub oder Krümel (auch von Tabak) solltest Du vorab mit einem weichen Tuch entfernen!

Und auch den Schutzumschlag brauchst Du beim Scannen nicht...

Die ersten Schritte

Wenn Du schon einen Scanner besitzt, wirst Du vermutlich im Grundsatz auch wissen, wie man ihn bedient. Und einiges hatte ich dazu ja auch schon in den vorhergehenden Kapiteln dieser Anleitung geschrieben. Dennoch möchte ich jetzt anhand der (manchmal etwas fuddeligen, weil unübersichtlichen) Software meines alten Flachbettscanners nochmal näher auf ein paar Punkte eingehen, die später der Texterkennung und auch Dir die Arbeit erleichtern werden, wenn es um das Erstellen von Ebooks aus gedruckten Texten geht.



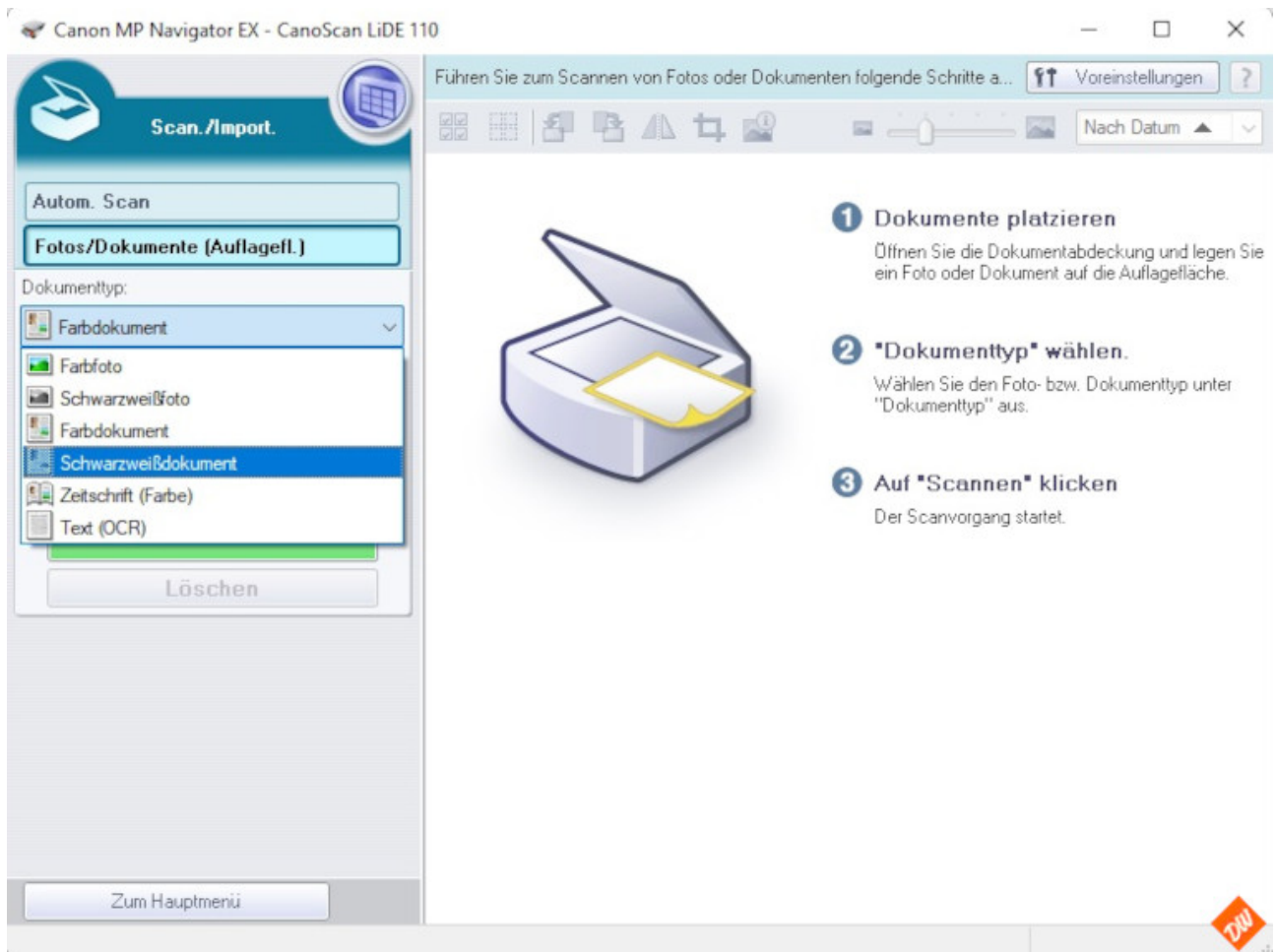
Wobei natürlich jedes Scannermenue anders aussieht und Du bei der Software Deines Scanners vermutlich manche Einstellungen an anderer Stelle findest.

Aber das macht auch nichts, denn prinzipiell sollten eigentlich alle Hersteller in ihrer Software alle Einstellungsmöglichkeiten bieten. Und zur Not hilft ja auch das Handbuch weiter.

Grundeinstellung .

Wichtig ist auf jeden Fall, von Anfang an den Modus zu wählen, der sich ja auch nach der weiter geplanten Verwendung der Scans richtet.

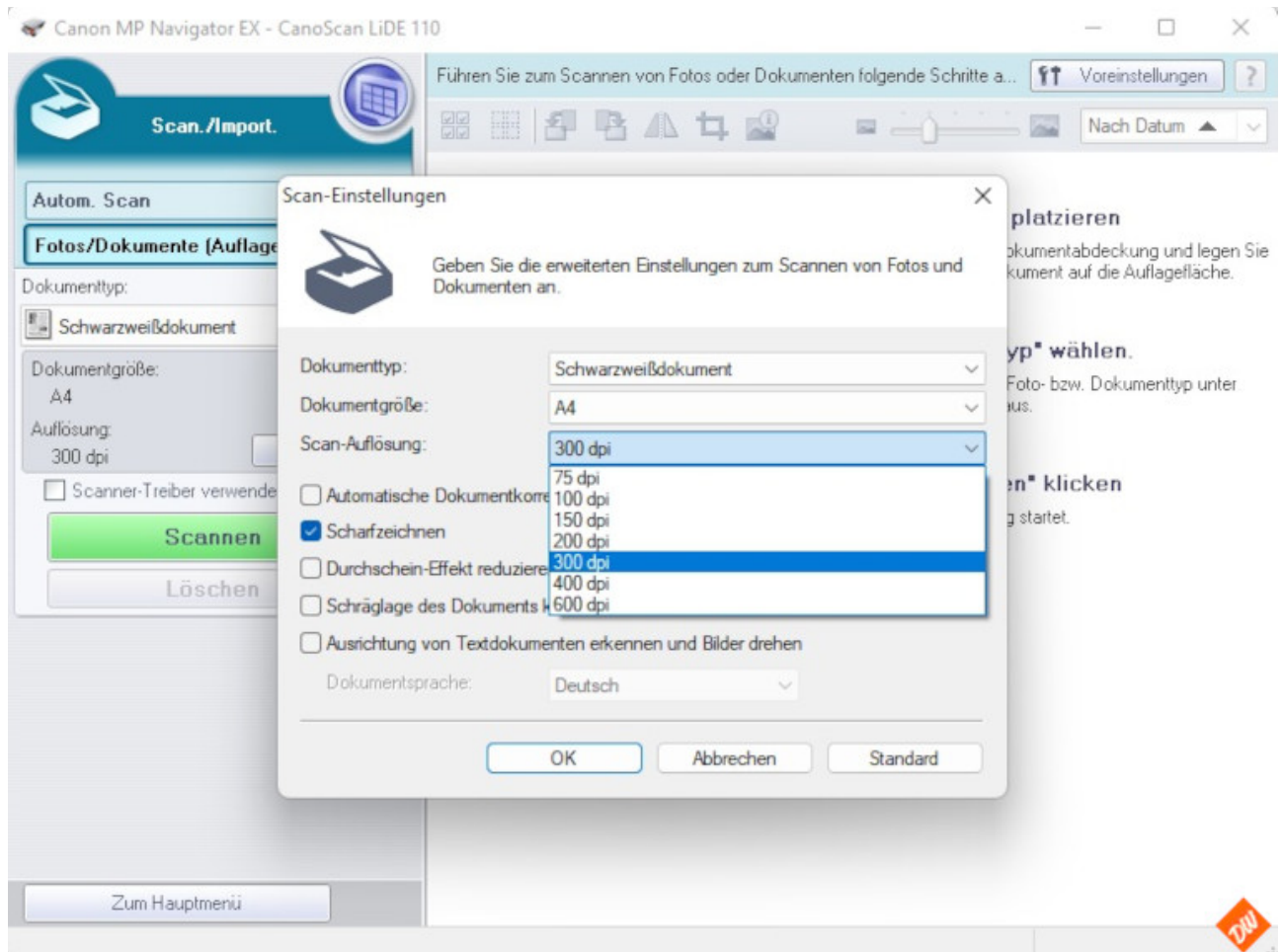
Möchte ich die Texte anschliessend weiter mit meiner Texterkennung bearbeiten, also auf jeden Fall „Schwarzweissdokument“ - möchte ich sie lediglich als „nur-PDF“ auf dem Tablett oder Rechner nutzen ist „Zeitschrift“ oder „Farbfoto“ die bessere Wahl.



Den auch an dieser Stelle schon angebotenen OCR-Modus können wir aber erst mal noch vernachlässigen, weil der für unser Zwecke kaum taugt und es zum erstellen eines Stichwortverzeichnisses für durchsuchbare „nur-PDF“ im späteren Verlauf noch eine weitere und bessere Variante gibt

Auflösung

Der zweite wichtige Punkt ist die Auswahl der richtigen DPI-Wertes, der für die Texterkennung am Besten 300 DPI betragen sollte, während für die Erstellung von „nur-PDF“-Formaten auch 150 DPI meist reichen



Dabei sind die 300 DPI ein Kompromisswert, zwischen Qualität, Dateigröße und Zeitaufwand. Höhere Auflösung könnten zwar ggf. die Ergebnisse der Texterkennung verbessern,

(erinnerst Du dich an die zu eng stehenden Worte, mit denen die Texterkennung bei allen meine Scannern Probleme hatte?)

bedingen aber auch, dass manche Scanner fühlbar langsamer arbeiten, der Speicherplatzbedarf für die Scan-Ergebnisse (egal ob JPG oder PDF) enorm steigt und die Texterkennung für ihre Arbeit noch mehr Zeit braucht - meist ohne das daraus grosse Gewinne für die nachfolgenden Arbeitsschritte entstehen würden:

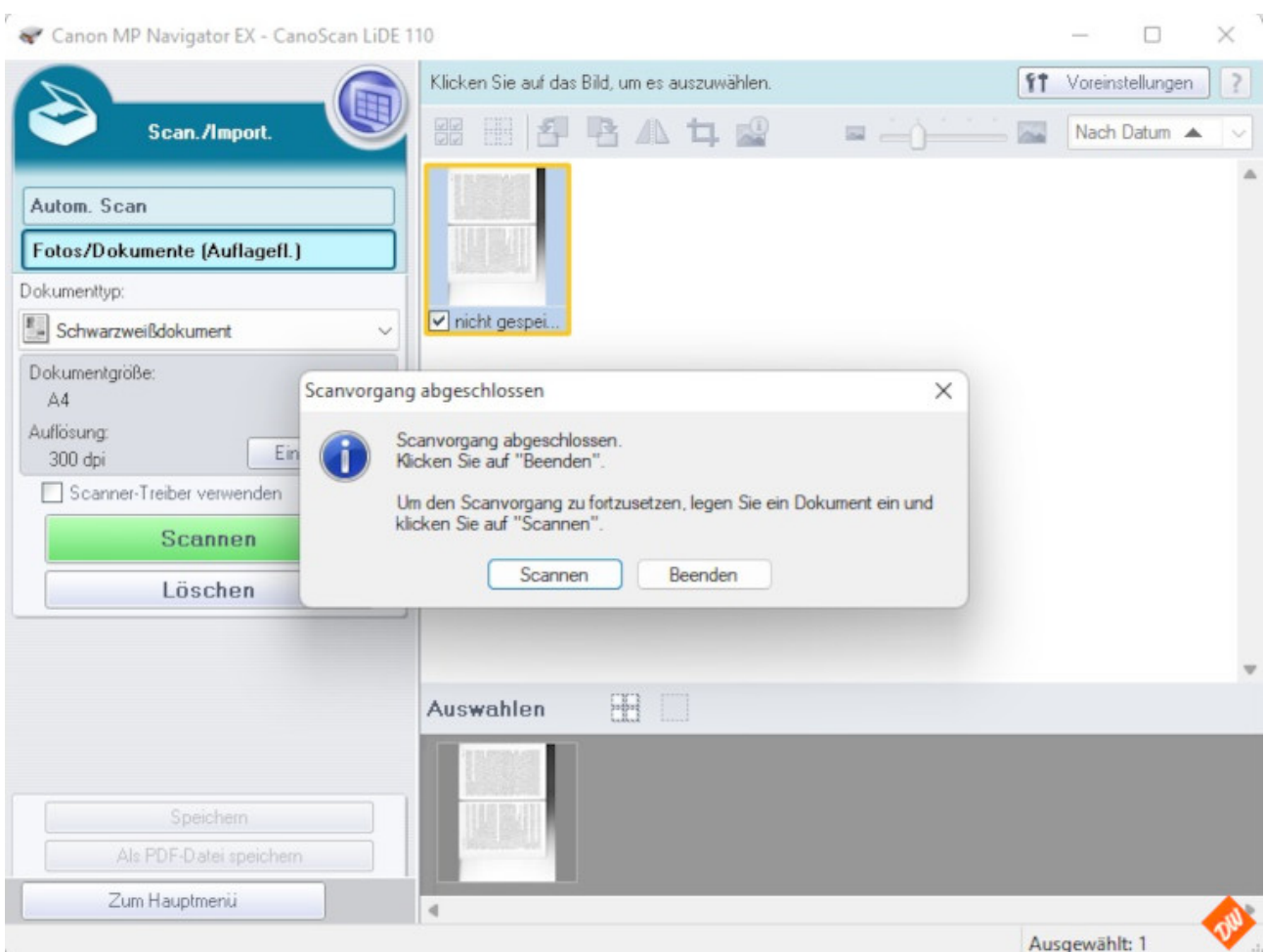
Mehrfach-Scans

Ich hatte ja schon erwähnt, das es aus meiner Erfahrung heraus am günstigsten ist, beim Scannen die Vorlage gleich in gut handhabbare Häppchen aufzuteilen.

20 bis maximal 40 Seiten sind dafür ein brauchbarer Wert, mit dem sowohl die Texterkennung als auch ich selbst bei der anschließenden Textverarbeitung gut klar kommen. Wobei ich tendenziell inzwischen meist bei Päckchen von 30 Seiten lande, je nachdem, wieviel Inhalt die einzelnen Seiten der Vorlage enthalten.

Ausnahmen davon mache ich nur, wenn ich Vorlagen habe, aus denen ein „nur-PDF“-Format entstehen soll. Dann scanne ich auch mal die ganze Vorlage am Stück oder Kapitelweise, wenn ich anschliessend ein Inhaltsverzeichnis brauche. (darauf ich aber im Thema Textverarbeitung auch nochmal näher ein)

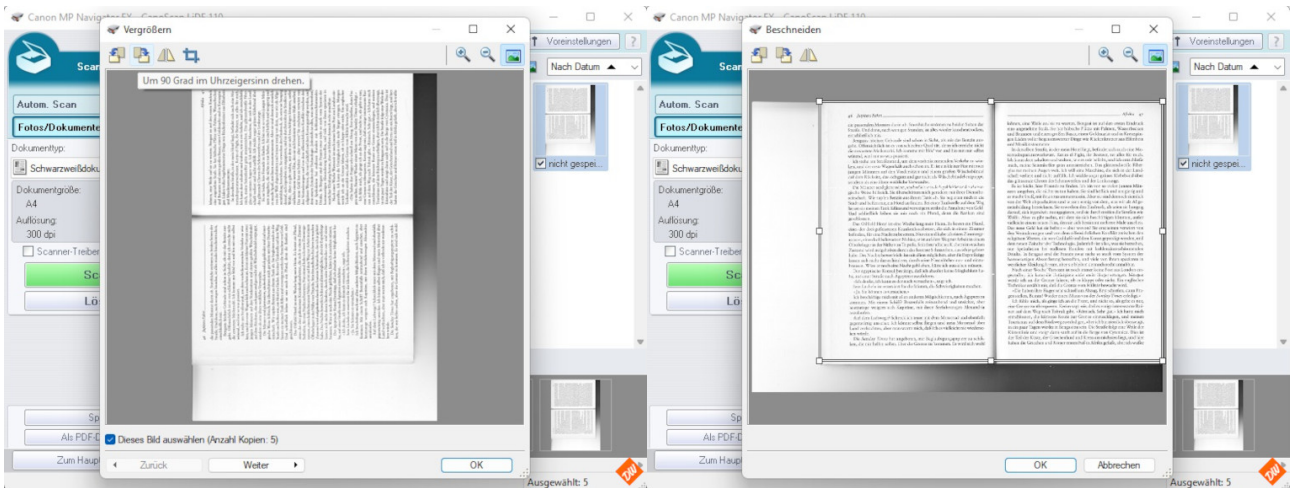
Netterweise bietet mir meine Scannersoftware für dieses Vorgehen auch gleich eine passende Möglichkeit:



Und ich vermute, diese Möglichkeit wird es auch bei anderen Herstellern geben – wenn auch womöglich an anderer Stelle innerhalb der Scannersteuerung.

Erste Bildbearbeitung

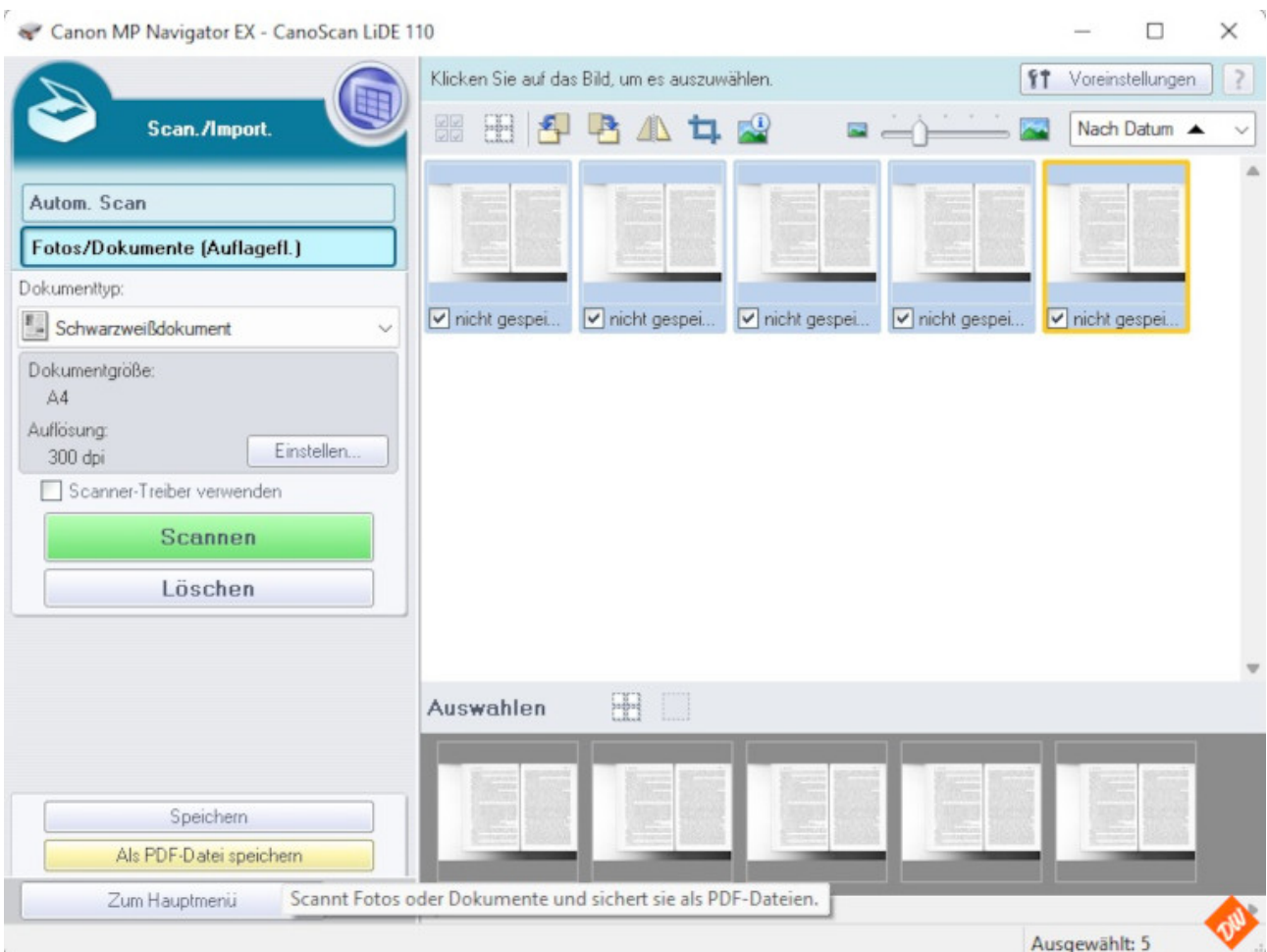
Während die Software meiner beiden anderen Scanner diesen Schritt gleich automatisch ausführt, muss ich beim Flachbettscanner vor dem Speichern noch selbst ein wenig Hand anlegen:

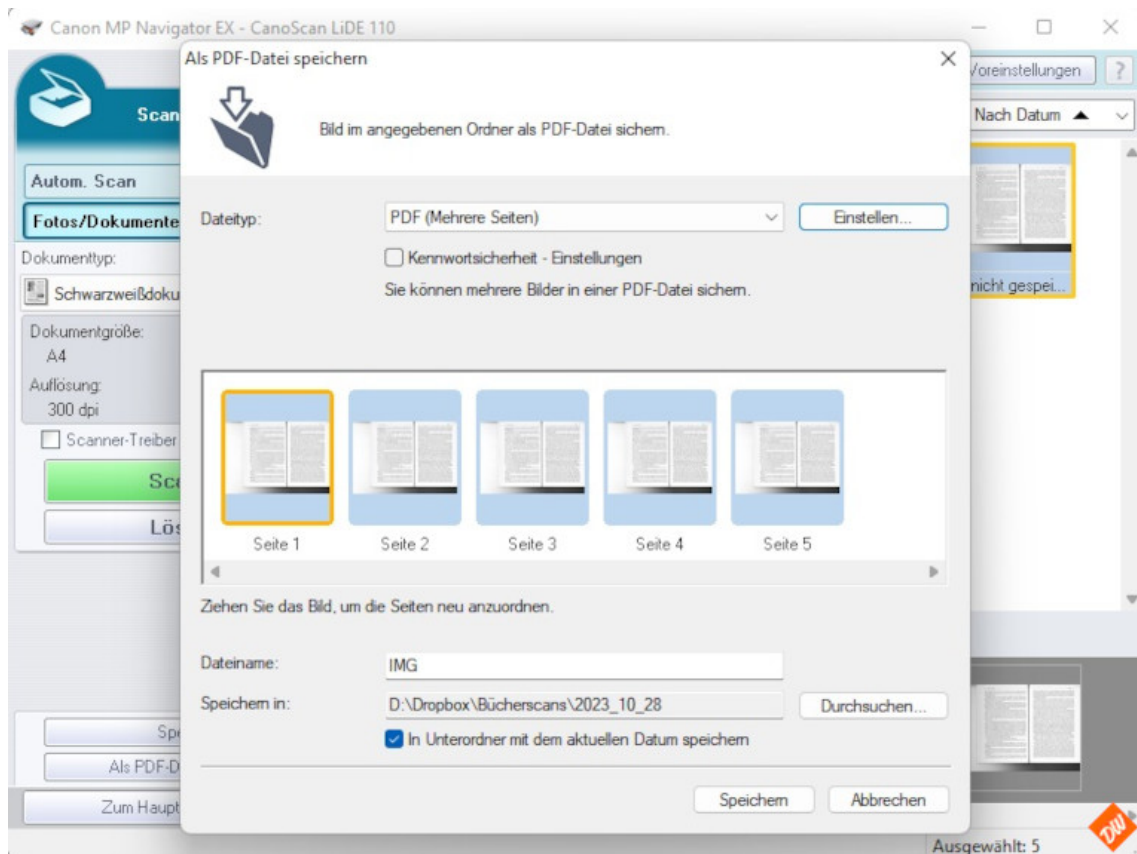


Ausrichten, weil Flachbettscanner prinzip-bedingt immer im Hochformat scannen und das bei einer Doppelseite natürlich nicht passt - und Zuschneiden auf den Inhaltsbereich als Vorbereitung für die Texterkennung – aber auch (bei „nur -PDF“) um die optische Anmutung der Scans ein wenig zu verbessern und gleichzeitig den Speicherplatzbedarf ein wenig zu verringern. Was nicht im Bild enthalten ist, frisst da auch keinen Platz.

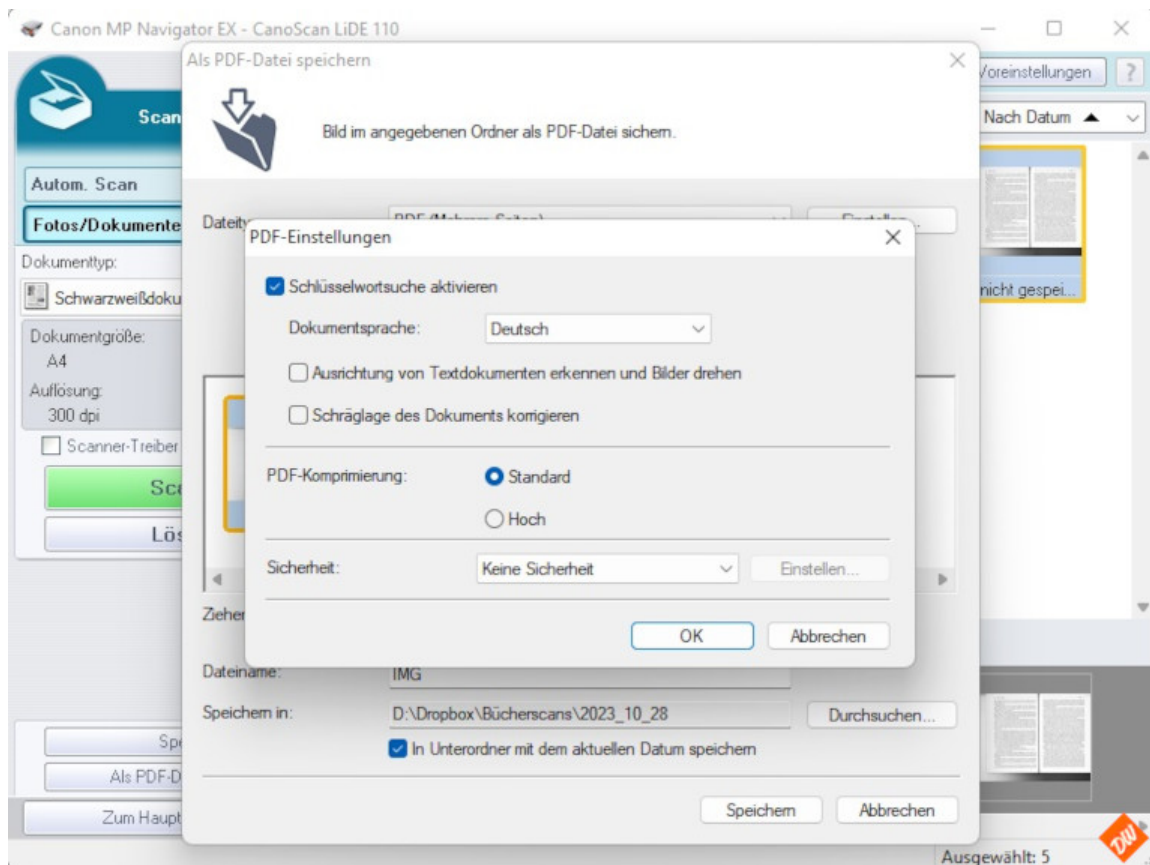
Speichern als PDF

Kommt als letzter Schritt das grosse Finale der Scan-Aktion: Der Export in eine PDF-Datei





und damit unter „Einstellungen“ auch die Möglichkeit, PDF-Dateien durchsuchbar zu machen, falls man das braucht – etwa für Hand-oder Fachbücher

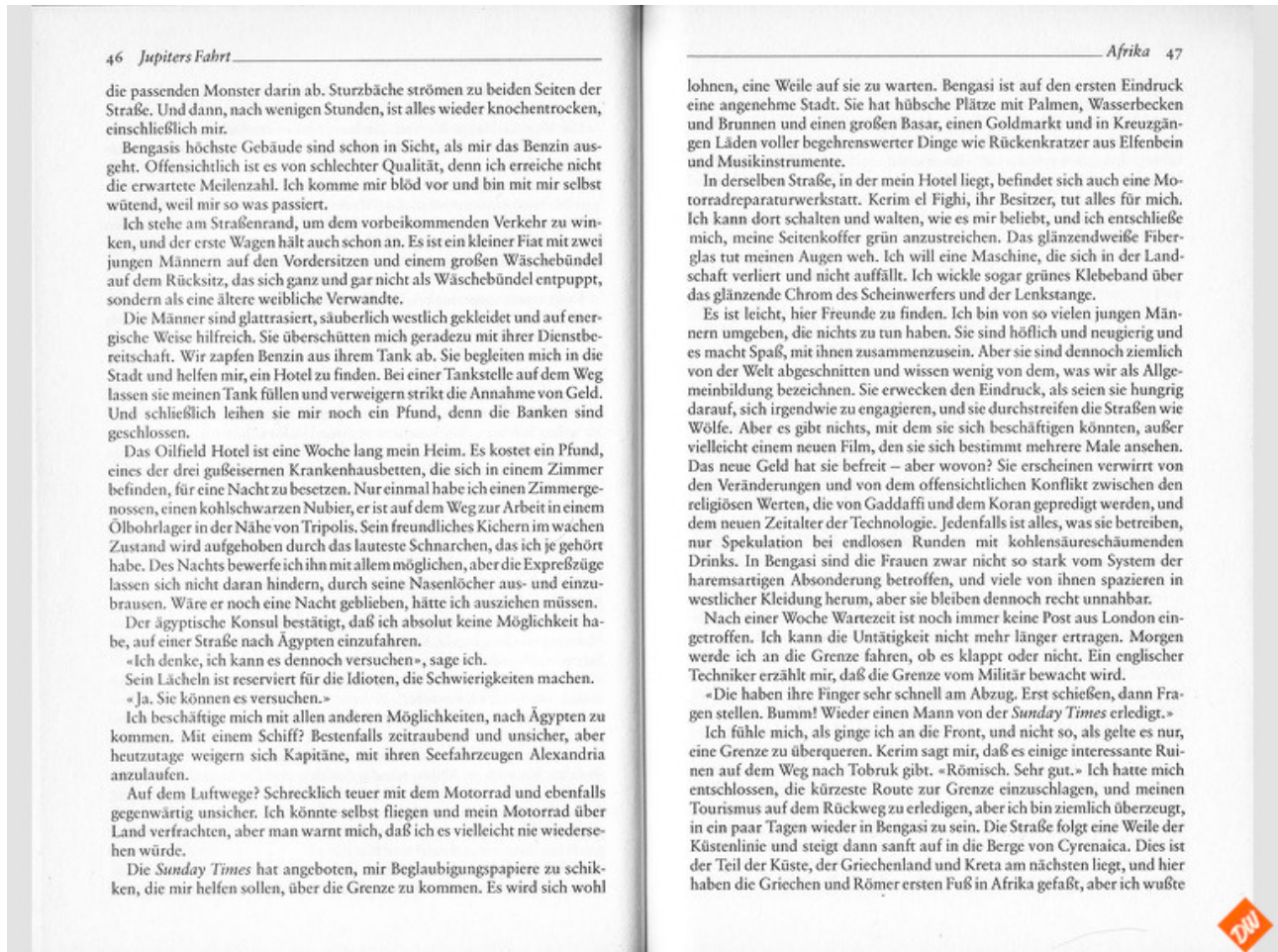


Doch das macht für unsere weitere Bearbeitung in der Texterkennung wenig Sinn und könnte sogar hinderlich werden, weil wir damit natürlich auch die PDF-Dateien noch weiter aufblähen.

Deswegen sollte man diese Option auf jeden Fall ausschalten, wenn man sie nicht braucht

Das Ergebnis

Damit kann man doch arbeiten!



Gute Scans minimieren nun mal die Fehlerquote bei der Texterkennung und damit auch die manuelle Nacharbeit bei der Textverarbeitung. Deshalb lohnt es sich auf jeden Fall, sich die Mühe sorgfältigen Arbeitens auch schon beim Scannen zu machen und dafür die notwendige Zeit zu investieren...