



Vom brüchigen Papier zur Textdatei

Minden-Lübbecker Kreisblatt und Mindener Tageblatt werden in Oberursel digitalisiert

Von Elke Büsching

Oberursel (mt). Berhane Mamas arbeitet konzentriert am Großscanner der Firma PPS PrePress Systeme in Oberursel. Dieser allerdings ist etwas ganz Besonderes. Eine Entwicklung des Firmenchefs Siegfried Peis, zum Scannen von Zeitungsseiten. Auch die vielen hunderttausend Seiten des Mindener Tageblatts, die sich seit 1856 angesammelt haben, werden hier eingelezen.

Mamas Blick gleitet von der alten MT-Doppelseite zu dem angeschlossenen Computer. Immer wieder müssen die Scanparameter neu eingestellt werden, um den unterschiedlichen Druckqualitäten gerecht zu werden. So beträgt der untere Seitenrand zuweilen meh-

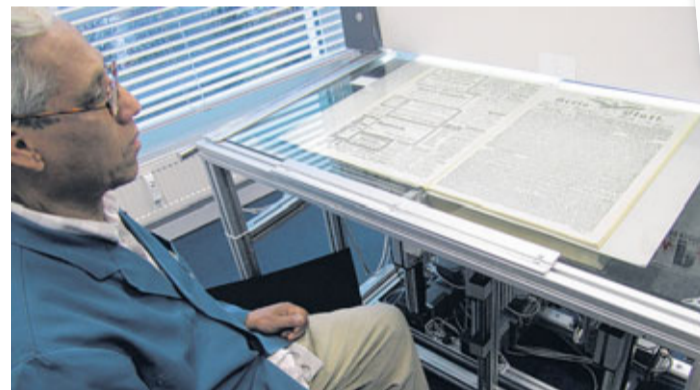
rere Zentimeter, dann wieder reicht die Schrift unmittelbar bis an das Seitenende heran. Ein kleiner Mittelsteg bildet die Seele des Scanners. Er bewirkt, dass die Zeitungsseiten vor dem Einlesen mittig geglättet werden. Auch die Seitenränder, die beim gewöhnlichen Scannen eines Buches meist verschwimmen, sind so einwandfrei lesbar.

Die Idee, Archive zu digital-

die alten Bände auseinander nehmen. Vorsichtig werden die alten Buchdeckel entfernt und die Nägel aus den dicken Büchern gezogen. Besonders behutsam geht er mit den alten Bänden des Mindener Tageblatts aus dem 19. Jahrhundert um. Das Papier ist brüchig und kann leicht zwischen den Fingern zerbröseln. Auch beim Einlegen der Seiten in den Scanner ist deshalb großes Fingerspitzengefühl gefragt.



Firmenchef Siegfried Peis (Foto li.) hat die Scanner entwickelt, an denen Berhane Mamas (Foto unten) die Seiten des Mindener Tageblatts einliest.



sieren ist gar nicht so neu. Bereits Anfang der 1990er Jahre gab es Software, mit der Archivgut verwaltet werden konnte. Das große Problem aber war: wie bekommt man die Archivalien in den Computer? Siegfried Peis erkannte die Notwendigkeit, einen Scanner zu entwickeln, der dies leisten konnte. „Letztlich unterscheiden sich die heutigen Scanner nicht wesentlich vom Prototyp, den ich damals skizziert und dann gebaut habe.“ Heute ist seine Firma mit zehn Mitarbeitern in der Lage, 250000 Seiten pro Monat zu verarbeiten.

Papier kann leicht zerbröseln

Bevor der Scanner zum Einsatz kommen kann, muss Buchbinder Sieghard Kleine

Die eingesetzten Scanner erlauben sowohl Graustufen-, Farb- und Schwarz-Weiß-Scans. Für die Digitalisierung von Tageszeitungen setzt Peis auf das bitonale, also Schwarz-Weiß-Scannen. Deutlich wird der Vorteil am Beispiel einer MT-Seite von 1928. „Hier hatte wahrscheinlich ein Druckerlehrling die Satzspalten nicht ordentlich gereinigt“, vermutet Peis. Die Seite weist große Unterschiede in der Farbtintensität auf, ist stellenweise kaum lesbar. Für die eingesetzten Scanner ist das kein großes Problem. Auch bei Farbdurchschlag, der bei dünnem Zeitungspapier häufig vorkam, werden mit dem bitonalen Scan sehr gute Ergebnisse erzielt. Einmal eingelezen, kommt eine intelligente Bildbearbeitungssoftware zum Einsatz, die Farbflecken und



Schmutzpartikel von den Seiten entfernt. Natürlich ohne Satzzeichen, i- und Rasterpunkte anzugreifen. Das beeindruckende Ergebnis ist eine einwandfrei lesbare Zeitungsseite mit reinweißem Hintergrund.

Softwareentwicklung als permanenter Prozess

Überhaupt schlägt nach dem Einscannen erst die Stunde der IT-Tüftler. Denn ohne Erkennungs- und Lesesoftware wären die alten MT-Seiten weiterhin nicht erschließbar. Dies ist die Domäne von Hans-Jürgen Döring und Arthur Bleile. Zunächst wird überprüft, ob auch wirklich alle Seiten eines Jahrgangs gescannt wurden. Dauerte dieser Vorgang früher einige Stunden, so hat Döring sich inzwischen eigene Tools geschaffen, die diesen Schritt in wenigen Sekunden erledigt haben.

Dann starten diverse Texterkennungs- und Korrekturprogramme für die Layout-Erkennung. Sie schaffen die Voraussetzungen für die automatische Artikel-Separierung. Artikel, die früher oftmals auf der nächsten Seite fortgesetzt wurden, müssen als zusammenge-

hörig erkannt werden. Auch die korrekte Zuordnung von Texten, Überschriften und Bildern stellt eine Herausforderung dar. Schließlich müssen die Seiten für die Nutzer aufbereitet werden. Eine Volltextsuche, die Eingrenzung nach Orten und Daten, die Suche nach Ressorts oder Autoren etwa werden ermöglicht. „Die Entwicklung dieser Software hat 25 Mannjahre gekostet“ weiß Chef Peis zu berichten. „Und sie ist nie abgeschlossen. Vielmehr ist dies ein lebender Entwicklungsprozess.“

Mit jedem Projekt tauchen neue Herausforderungen auf, für die Lösungen gefunden werden. So ist das Scannen der ersten Jahrzehnte des Mindener Tageblatts – damals noch Minden-Lübbecker Kreisblatt – ein Novum für PPS PrePress Systeme: Älter war bislang keine Zeitung, die hier gescannt wurde. Entsprechend anspruchsvoll ist die Aufbereitung der alten Frakturschrift. Für die Tüftler aus Oberursel ist das aber allemal ein lösbares Problem.

Sieghard Kleine (Foto oben) muss zunächst die alten Buchdeckel entfernen, bevor mit dem Einscannen der Seiten begonnen werden kann. Anschließend bekommen die historischen Bände eine neue Bindung. MT-Fotos: Manfred Otto/Elke Büsching/pr

