

Heider-Tauf- und Sterberegister: Benutzeranleitung

Benutzeranleitung zur OCR-Umwandlung, Formatierung und Prüfung der Heider-Register

- 295 -

Tag	Name	Eltern, Beruf, Ort	Tom.
5.12.1670	Stockmayer Nikolaus	Thomas - Martha Niederpuemberg	I 159
20. 3.1671	Steininger Joachim	Johann - Christina zu Püreth	I 162
7. 9.1671	Stockinger Tobias	Jakob - Eva an der Donau	I 163
6. 5.1672	Stockmayer Tobias	Tobias - A. Maria zu Harbach	I 167
17. 8.1672	Stöger Regina	Johann - Eva an der Lohrmühl	I 168
16. 9.1672	Steininger Elisabeth	Johann - Christina zu Püreth	I 169
28. 9.1672	Stelzer Martha	Adam - Maria Pruckmühl	I 169
29.11.1673	Steininger Andreas	Johann - Christina zu Püreth	I 173
20. 7.1674	Stockmayer Andreas	Andreas - A. Maria zu Harbach	I 175
19. 3.1675	Stöger Josef	Johann - Eva an der Lohrmühl	I 177
9. 7.1675	Stöckl Heinrich	Georg - Eva	I 178
11. 3.1676	Stockmayer Magdalena	Stefan - Katharina	I 182
11. 3.1677	Stockmayer Josef	Tobias - Maria	I 185
3. 8.1677	Stöger Clara	Johann - Eva an der Lohrmühl	I 187
5. 2.1678	Stöckl Mathias	Georg - Eva	II 1
31. 8.1679	Stockmayer Adam	Tobias - Maria	II 10
1. 9.1680	Stöckl Rosina	Georg - Eva	II 17
22.10.1680	Stöger Barbara	Johann - Eva	II 17
3. 9.1682	Stockmayer Nikolaus	Tobias - Anna Maria	II 29
28. 7.1683	Stöger Stefan	Johann - Eva	II 36
8. 9.1683	Stöckl Maria	Georg - Eva am Stöckl	II 36
26. 1.1684	Strasser Euphrosina	Wolfgang - Sara Wizersdorf	II 38
17. 5.1684	Stockmayer Anton	Tobias - Anna Maria zu Harbach	II 40
26. 5.1684	Starnberger Justina	Joachim - Sophia Wizersdorf	II 40
9. 2.1656	Strasser Susanna	Wolfgang - Sara Riezersdorf	II 49
21. 4.1686	Starnberger Katharina	Joachim - Sophia Wizersdorf	II 50

Verfügbare Dokumentationen zu den Themen Kirchenmatriken, Digitalisierungsprojekt der Heider-Register und Josef Heider auf den Netzseiten der Familia Austria

- *Lebenslauf von Prof. Josef Karl Heider*
mit einleitenden Gedanken zur Familienforschung und zwei Fotos, zur Verfügung gestellt von seinen Erben
Autor: Josef Heider
- *Einführung zu den Matrikeln*
Autor: Josef Heider, Wien 1983
- *Verzeichnis der Register zu den Matrikeln der Pfarren des Mühlviertels und des Salzkammergutes*
mit Angaben über den Inhalt der einzelnen Bände.
Autor: Josef Heider, Wien 1983
- *Heider-Register und Digitalisierungsprojekt: Einführung*
Was sind die Heider-Register und wozu dient ihre Digitalisierung durch den Verein Familia Austria zusammen mit dem Oberösterreichischen Landesarchiv?
Autor: Sepp Asanger
- *Heider Tauf- und Sterberegister: Benutzeranleitung zur Datenerfassung*
Benutzeranleitung zum Einrichten der OCR-Software und des PC-Editors KEDIT, zur OCR-Umwandlung, Formatierung und Prüfung der Heider-Register
Autor: Sepp Asanger
- *Heider-Projekt: technische Dokumentation*
beschreibt technische Grundlagen, Dateiverwaltung, Namensschemata, Nachbearbeitung von Scan-Dateien
Autor: Sepp Asanger



INHALTSVERZEICHNIS

1	Einleitung	6
2	Die Tauf- und die Sterberegister von Josef Heider	7
2.1	Aufbau der Taufregister	7
2.2	Aufbau der Sterberegister	8
2.3	Erklärungen zu den vier Spalten Tag, Name, Eltern... und Tom.....	9
2.3.1	Tag-Spalte	9
2.3.2	Namens-Spalte	9
2.3.3	Spalte Eltern. Beruf, Ort (EBO).....	9
2.3.4	Tom.-Spalte	9
2.4	Verarbeiten der Registerdaten in zwei Verarbeitungsschritten	10
3	OCR-Schrifterkennung mit ABBYY FineReader	11
3.1	Verwendete Software.....	11
3.2	Einrichten des FineReaders	11
3.3	Öffnen der zu übersetzenden Scans (Registerseiten).....	11
3.4	Bereich der Texterkennung.....	12
3.5	OCR-Erkennung	14
3.6	Arbeiten mit den umgewandelten Registerseiten.....	14
3.6.1	Zoom-Fenster	14
3.6.2	blau markierte Zeichen, die nicht korrigiert werden müssen	15
3.6.3	Manuelle Prüfung der umgewandelten Texte	16
3.6.4	Prüfung von Namen und Begriffe mit dem Wörterbuch von FineReader	16
3.6.5	Leerzeichen	16
3.6.6	Trennung von Begriffen (Namen, Beruf, Ort) auf zwei aufeinanderfolgende Zeilen.....	17
3.6.7	Bearbeitung rückgängig machen	17
3.6.8	Speichern der bearbeiteten Text-Datei (Registerseite).....	17
3.6.9	Unterbrechung der OCR-Bearbeitung	18
4	Formatierung und Datenprüfung der Registerseiten mit dem Texteditor KEDIT	19
4.1	KEDIT	19
4.2	KEDIT-Profile	19
4.2.1	Startposition der 4 Spalten (Blöcke).....	20
4.2.2	Reihenfolge von tomus und pagina	20
4.2.3	Liste der tomus-Nummern	20
4.3	Aufruf von Funktionen für die Tauf- und Sterberegister über Funktionstasten.....	20
4.4	Starten der Formatierung und Prüfung einer Registerseite	21
4.5	Lesen der vorhergehenden Seite.....	21
4.6	Seitennummer	21
4.7	Überschriftzeile	22

4.8	Leerzeilen	22
4.9	Eintragung	22
4.9.1	Name	22
4.9.2	Vertauschte Vor- und Familiennamen	23
4.9.3	Sonderfälle beim Namen.....	23
4.10	Eltern, Beruf, Ort (EBO).....	24
4.11	Zahlenangaben in der EBO-Spalte.....	24
4.12	tomus-Block.....	24
4.12.1	tomus-Block beim Taufregister	24
4.12.2	tomus-Block und Altersangabe beim Sterberegister	25
4.13	falsche Zeilen	28
4.14	Manuelle Korrekturen in den Heider-Registern durch Benutzer	29
4.15	Blockzeile in der Text-Datei	29
4.16	Zwillinge und mehrfache Vornamen	30
4.17	Fehlermeldungen, Kontrolle.....	31
4.18	Nachtrag	32
4.19	Rückgängig machen von Änderungen	32
4.20	Anhang	33
5	OCR-Schrifterkennung mit ABBYY FineReader.....	34
5.1	Die FineReader-Optionen zum Einrichten und Anpassen	34
5.1.1	Dokument	34
5.1.2	Scannen/Öffnen.....	34
5.1.3	Lesen: Gründlich, Benutzermuster	35
5.1.4	Speichern: TXT	35
5.1.5	Anzeigen: Schriftart für Text	36
5.1.6	Erweitert.....	37
5.1.7	Zwischenspeichern der OCR-Ergebnisse	37
5.2	Benutzerwörterbuch von FineReader.....	37
5.2.1	Benutzerwörterbuch als wichtiges Hilfsmittel zur Datenprüfung.....	37
5.2.2	Benutzerwörterbuch erstellen.....	38
5.2.3	Neue Wörter im Benutzerwörterbuch speichern	39
5.2.4	Benutzerwörterbuch exportieren/importieren.....	39
6	KEDIT Texteditor: Installation und Editierfunktionen.....	41
6.1	Installation der KEDIT-Software und der Verarbeitungs-Programme (Macros)	41
6.2	Die wichtigsten Editor Zeilenbefehle	41
6.2.1	Einfache Befehle	41
6.2.2	Befehle mit Zahlenangaben.....	41
6.2.3	Verdoppelung der Befehle	42
6.3	Arbeiten mit Textblöcken.....	42

6.3.1	Textblock in den Kommentar verschieben	43
6.3.2	Verschieben der Altersangaben in den Sterberegistern	43
7	Abbildungsverzeichnis	45

1 Einleitung

In diesem Dokument geht es nicht darum, das Heider-Projekt insgesamt zu beschreiben. Die allgemeine Beschreibung der Heider-Register, des Digitalisierungsprojektes und den gesamten Ablauf kann man in der Dokumentation *Heider-Register und Digitalisierungsprojekt Einführung* auf den FA Netzseiten nachlesen. Diese Doku ist eine Anleitung, wie die gescannten Tauf- und Sterberegister von Josef Heider in Textdateien umgewandelt, durch Programme in eine einheitliche Form gebracht und geprüft und wie die Daten kontrolliert werden und für die weitere Bearbeitung aufzubereiten sind. Dabei ist auf formale Dinge zu achten, damit die nachfolgende Aufteilung der Daten in die Spalten der Geburtsdatenbank mit den von Dr. Peter Haas entwickelten Programmen rascher funktioniert. In eigenen Kapiteln gibt es Hinweise zum Einrichten der erforderlichen Software-Komponenten ABBYY Fine Reader und KEDIT. Da die Erfassung der Heirats-Register bereits 2019 abgeschlossen wurde, wird auf die besonders komplexen Register in diesem Dokument nicht mehr eingegangen.

Die Beschreibung geht davon aus, dass die Registerseiten von den Heider-Büchern bereits eingescannt und gerade ausgerichtet sind, die Einzeldateien richtig durchnummeriert und in einem Scan-Ordner gespeichert sind. Informationen dazu gibt es im Dokument *Heider-Projekt: technische Dokumentation*.

2 Die Tauf- und die Sterberegister von Josef Heider

Den allgemeinen Aufbau der Heider-Register und Hinweise zum Alphabet und zum Schriftbild ist im Dokument *Heider-Register und Digitalisierungsprojekt Einführung* beschrieben. In den folgenden Kapiteln gibt es kurze Wiederholungen und zu beachtende Sonderheiten mit Beispielen.

2.1 Aufbau der Taufregister

- 36 -

Tag	Name	Eltern, Beruf, Ort	Tom.
17. 9.1720	Pesenböck Katharina	Simon - Eva Ebretschlag	II 370
7.11.1720	Pamer Elisabeth	Johann - Anna Maria Amessschlag	II 371
7.11.1720	Pachner Leopold	Michael - Katharina Bürger	II 371
27. 2.1721	Prieschl Elisabeth	Michael - Rosina Ebretschlag	II 373
19. 3.1721	Pachner Maria Josefa	Joh.Michael - A.Magdal. Ratsbürger	II 374
29. 3.1721	Pesenböck Anna	Wolfgang - Sophia Amessschlag	II 375
8. 5.1721	Pichler Franz	Josef - Elisabeth Bürger	II 376
27. 5.1721	Pichler Johann	Andreas - Maria Oberstiftung	II 377
7. 9.1721	Pernauer Adam Heinr.&	Joh.Mich. Josef-Kath. Bürger	II 379
9.10.1721	Pogner Martin	Georg - Apollonia Silberhartschlag	II 379
13.10.1721	Brungraber Leopold	Thomas - Salome Bürger	II 379
16.11.1721	Proier Maria	Mathias - Maria Langbruck	II 380
19. 1.1722	Prandstetter A.Maria	Thomas - Susanna	II 382

Abb. 1: Seite eines Taufregisters

Zum Verständnis sei zunächst das Format der Heider-Register erklärt. Im Beispiel sehen wir den grundsätzlichen Aufbau einer Registerseite für Taufen mit einer Seitennummer oben zwischen zwei – Zeichen. Die Seitennummer kann in der Mitte oder am rechten Rand stehen. Nach ein bis zwei Leerzeilen kommt zwischen 2 Querlinien die Spaltenüberschrift und danach beginnen die Registereintragungen in 4 Spalten (Blöcken):

- **Tag** Datum der Taufe (Geburt) bzw. Sterbedatum
- **Name** Familien- und Vorname(n) des Kindes bzw. der Kinder oder des Verstorbenen
- **Eltern, Beruf, Ort (EBO)**, gemischte Angaben zu den Eltern
- **Tom.** Buchnummer (tomus) und Seitennummer (pagina) des Kirchenbuches

Heider hat seine Register einseitig auf sehr dünnem Papier mit Schreibmaschine geschrieben und war dabei auf eine Seitenbreite von etwas weniger als dem A4-Format beschränkt. Bei längeren Namen wurde die Einteilung in die 4 Spalten daher nicht immer eingehalten. Bei den Zwillingen am 7.9.1721 ist zu sehen, wie die Vornamen der Kinder in die Elternspalte hineinragen. Beim manuellen Lesen ist das zwar meistens klar erkennbar, doch die automatische Erkennung der Spalten durch ein Standardprogramm am PC versagt, weil dort eine fixe Spalteneinteilung Voraussetzung ist.

Weitere Einsparungen im Platzbedarf hat Heider durch Weglassung von trennenden Leerzeichen erreicht. Wir sehen das bereits bei den erwähnten Zwillingen (Adam Heinr.& Joh.Mich.), als auch beim Bindestrich zwischen den Vornamen der Eltern. Bei der Eintragung vom 19.3.1721 fehlt das Leerzeichen sowohl bei den Vornamen der Eltern, als auch als auch zwischen der Mutter und dem tomus (A.Magdal.II 374).

2.2 Aufbau der Sterberegister

- 2 -

Tag	Name	Eltern, Beruf, Ort	Tom.
18.10.1674	Aichinger Stefan	Aigen	II 247
20. 8.1674	Autengruber Michael	Rudolfing	II 247
13. 5.1675	Autengruber Anna		II 248
11. 9.1675	Autengruber Martha	Totengräber, Wächter	II 248
4. 6.1678	Amprosin Maria	Praitenstein	II 254
6. 6.1682	Almesberger Kind	Florian, Schindlau	II 261
23. 3.1684	Autengruber Andreas	Rudolfing	II 265
2.12.1688	Almesberger Adam	Aigen	II 274
14. 9.1715	Almesberger Kaspar	Schindlau	6o III 496
22. 6.1716	Almesberger M.Magdalena	Schindlau	5 III 498
9. 6.1718	Almesberger Maria	Wurmbrand	6 III 5o5
6.12.1719	Asanger Johann	Knecht zu Aigen	3o III 5o9
5. 5.1723	Aigner Lukas	Landgerichtsdieners zu Aigen	6o IV 589
24. 1.1725	Almesmüller Philipp	Wurmbrand	7o IV 595
26. 8.1726	Almesberger Magdalena	Hans Georg, Schindlau	IV 6o3 15M
9. 7.1787	Aigner Hans Georg	ledig, Aigen	28 IV 6o9
6. 2.1729	Asanger Maria	Michael, ledig zu Unterneudorf	11W IV 613
7. 4.1729	Almesberger Gertrude	Hans Georg	7 IV 615
22.12.1729	Asanger Margarete	Holzhauseln	6o IV 619
10. 3.173o	Aigner Eva Maria	Landgerichtsdieners zu Aigen	71 IV 621
17. 9.173o	Aigner M. Katharina	Kaspar, Langerichts- diener zu Aigen	7T IV 624
9. 7.1731	Andexlinger Maria	Rudolfing	51 IV 627
29. 9.1731	Aigner M. Magdalena	Kaspar, Landgerichts- diener zu Aigen	IV 628 low
27.12.1731	Aigner Anna Maria	Landgerichtsdieners zu Aigen	24 IV 629

Abb. 2: Seite eines Sterberegisters

Die Sterberegister haben denselben grundsätzlichen Aufbau, aber eine neue Information kommt hinzu: das Alter. Die besondere Schwierigkeit besteht dabei darin, dass die Altersangabe in keiner eigenen Spalte steht. Sie steht zwar vor dem tomus, aber am Ende der Elternspalte und ist damit Teil derselben. Außerdem ist sie optional und fehlt sehr häufig. Zwingend ist damit nur die tomus-Angabe (und selbst die fehlt manchmal, wenn der Platz zu eng wurde). Als ‚Ankerpunkt‘ vor allem bei der Erstformatierung dient daher die Position von Tomus. Um Platz zu sparen, hat Heider die Altersangabe manchmal in die EBO-Spalte der Folgezeile geschrieben. In obigem Beispiel sehen wir das bei den Eintragungen vom 26.8.1726 und vom 29.9.1731. Wir sind hier also mit einigen Unsicherheiten konfrontiert

2.3 Erklärungen zu den vier Spalten Tag, Name, Eltern... und Tom.

2.3.1 Tag-Spalte

Form und Inhalt der Spalten können stark variieren. So steht etwa in Spalte **Tag** in der Regel ein gültiges Datum, manchmal aber fehlen Tag und/oder Monat. Auch Angaben wie Ostern oder Fasching statt des konkreten Datums kommen vor.

Das Format des Datums ist im Normalfall **tt.mm.jjjj** und beginnt Stelle (column) 1, wenn der Tag 2-stellig ist. Ein einstelliger Tag beginnt auf der Stelle 2. Es kommt aber immer wieder vor, dass auch der einstellige Tag auf der Stell 1 beginnt. Das muss durch Einfügen einer Leerstelle korrigiert werden (siehe [Abb. 4: Auswahlrahmen erstellen](#) auf Seite 13).

2.3.2 Namens-Spalte

Die Namens-Spalte ist relativ einheitlich, nur manchmal sind Familien- und Vornamen vertauscht. Die Namen müssen dann manuell umgedreht werden. Vornamen von Mehrlingen sind durch das **&**-Zeichen gekennzeichnet bzw. verbunden. Vor allem bei den seltenen Drillingen (Vierlinge sind mir bisher nur einmal untergekommen) reichen die Namen weit in die nächste Spalte, die Elternspalte, hinein.

2.3.3 Spalte Eltern. Beruf, Ort (EBO)

Diese Spalte kann sehr gemischte Informationen enthalten, die manchmal für den Leser, aber noch viel mehr von einem Programm, nicht einfach zu interpretieren sind. Die Vornamen der Eltern sind mit einem Bindestrich verbunden, vor und nach dem eine Leerstelle sein muss (im Unterschied zu verbundenen Namen). Bei Platznot wurden aber oft die Leerstellen weggelassen: *Ratsbürger, Fleischhacker*. Die Begriffe können auch abgekürzt oder abgeteilt und in der nächsten Zeile fortgesetzt sein:

*Witwer, Gemainbraumei-
ster*

Schwierig wird die Interpretation vor allem bei ledigen Elternteilen, die durch die Zusätze **ledig, ill.** oder **fil. ill.** gekennzeichnet werden (=Familienstand). Nach **fil. ill.** folgen z.B. oft Namen der Großeltern(-teile) und fallweise deren Ort und/oder Beruf des Großvaters. Eine andere Form der Angabe von ledigen Eltern ist die Kennzeichnung mit **V. :** bzw. **M. :**. Hier wird, nachdem der eine Elternteil beschrieben wurde, der zweite angeführt, manchmal mit dem Vermerk **verheiratet**, also ein außereheliches Kind.

Es gibt dann noch andere Vermerke wie **Soldat**, meist mit Angabe des Regimentes, **Findelkind** und mehr.

2.3.4 Tom.-Spalte

In der Spalte **Tom.** stehen die Nummer des Kirchenbuches (tomus) und die Seitennummer. In dieser Beschreibung verwende ich den Begriff **pagina** oder **pag** für die Seitennummer des Kirchenbuches, um sie von der Seitennummer der Heider-Register zu unterscheiden.

Josef Heider hat in seinem Verzeichnis der von ihm registrierten Kirchenbücher die Originalbezeichnung der Kirchenbücher und den Zeitbereich tabellarisch zusammengestellt. In der Regel hat er diese Originalbezeichnung in den Registern übernommen, aber nicht immer. Z.B. sind bei den Sterbebüchern von Altmünster die Originalbücher mit den römischen Ziffern XIX bis XXIV bezeichnet. Der vorletzte Band, den Heider registriert hat, hat die Bezeichnung XXIII, also eine fünfstellige römische Nummer. Vermutlich um Platz zu sparen, hat Heider in den Sterberegistern statt der römischen Nummern lateinische verwendet. In matricula wiederum wurde ein eigenes Nummerierungssystem eingeführt, das weder mit den römischen, noch mit den lateinischen Nummern von Heider übereinstimmt. In unseren digitalisierten Registern haben wir die von Josef Heider-Register vergebenen tomus-Nummern unverändert zu übernommen. Tomus und pagina sollten innerhalb eines Indextabstaben aufsteigend sein. Eine krasse Abweichung von dieser Regel entsteht bei den Kirchenbüchern, die nach Ortschaften geführt wurden. Da beginnen innerhalb desselben Buchstabens die Matrikeln je Ortsteil datumsmäßig wieder von vorne. Dadurch springen die Seitennummern im Register, das ja nach dem Anfangsbuchstaben des Namens und nach Datum sortiert ist, ständig auf und ab und entziehen sich damit einer Kontrolle.

2.4 Verarbeiten der Registerdaten in zwei Verarbeitungsschritten

Das Verarbeiten der Registerdaten erfolgt in zwei voneinander unabhängigen Schritten:

- Optische Schrifterkennung (OCR) mit ABBY-Fine Reader,
- Formatierung und Datenprüfung mit KEDIT.

In jedem der beiden aufeinanderfolgenden Verarbeitungsschritte können die Registerdaten verändert bzw. korrigiert werden. Aufgrund der spezifischen Prüfungen und Rückmeldungen ergeben sich aber Schwerpunkte für das Eingreifen und Korrigieren.

Die OCR-Umwandlung markiert unsichere Zeichen mit blauer Farbe und unbekannte Wörter werden rot unterstrichen, daher sind diese erste Kandidaten zur Überprüfung. Andererseits erkennt KEDIT Probleme beim Datum, tomus und pagina und meldet dies mit einer Nachricht. Außerdem werden dort die 4 Spalten schon sauber getrennt, was die optische Prüfung erleichtert.

Manchmal hat J. Heider etwas ausführlichere Texte aus den Matrikeneinträgen übernommen. Diese stehen dann teilweise quer über die Namens- und EBO-Spalte und sollten als Kommentar übernommen werden. Dazu eignet sich am besten die Funktion Textblöcke im KEDIT. Siehe Beispiel *Textblock in den Kommentar verschieben* auf Seite 43.

3 OCR-Schrifterkennung mit ABBYY FineReader

3.1 Verwendete Software

Als OCR-Software (Optical Character Recognition, optische Buchstabenerkennung) verwenden wir den ABBYY FineReader 9.0 Professional Edition von Adobe. Der Verein Familia Austria hat für das Projekt 3 Lizenzen angeschafft. Kauf am 20.6.2011 von CH.S-Shop, Kunden-Nr.: 179729878, Serien-Nr. FPEF-9010-0001-3425-7224-7882. Davon verwende ich eine auf meinem PC und eine auf dem Laptop. Die 3. Installation hat meines Wissens Dr. Peter Haas. Eine von diesen 3 Lizenzen müsste deaktiviert (deinstalliert?) werden, um sie für einen neuen User frei zu bekommen.

Ende 2014 habe ich die Version 12 versuchsweise probiert. Die bestehenden Probleme mit Version 9.0 (Wörterbuch) wurden aber nicht verbessert. Außerdem ergab sich ein neues Problem, weil einstellige Leerzeichen manchmal durch eine Leerstelle (wie bisher), manchmal aber auch durch ein Tabulatorzeichen ersetzt werden. Damit verliert eine Textzeile völlig seine Form und das Ergebnis war unbrauchbar. Das Problem konnte damals mit der Supportstelle von Abbyy weder erklärt noch gelöst werden.

FineReader versucht, aus der Bilddatei (die digitalisierten Buchseiten sind computertechnisch nur Bilder) die Texte zu erkennen und damit für den Computer lesbar zu machen. Aus den Bilddateien entstehen dadurch Textdateien, die die Grundlage für die weitere Verarbeitung sind. Die Erkennungsquote der Texterkennung hängt wesentlich von der Schrift- und von der Scanqualität der Registerseite ab. Die Scan-Qualität kann vor allem bei sehr hellen Schriften durch Erhöhung des Kontrastes verbessert werden, was zu einer ganz entscheidenden Verbesserung der Texterkennung führen kann. Ich benütze dazu die Fotobearbeitung in meinem Foto-Programm und habe damit zu einer mehrfachen Verbesserung der Texterkennung erreicht. Siehe auch das Kapitel *Erfassen und Nachbearbeiten der gescannten Registerseiten* im Dokument *Heider-Projekt technische Dokumentation*.

3.2 Einrichten des FineReaders

Für ein erfolgreiches Arbeiten muss ABBYY FineReader an die speziellen Bedürfnisse dieses Projektes angepasst werden. Dies geschieht vor allem mit der Funktion *Extras/Optionen...* Das Einrichten des FineReader wird in dem Kapitel, *5.1 OCR-Schrifterkennung mit ABBYY FineReader* auf Seite 34 beschrieben.

3.3 Öffnen der zu übersetzenden Scans (Registerseiten)

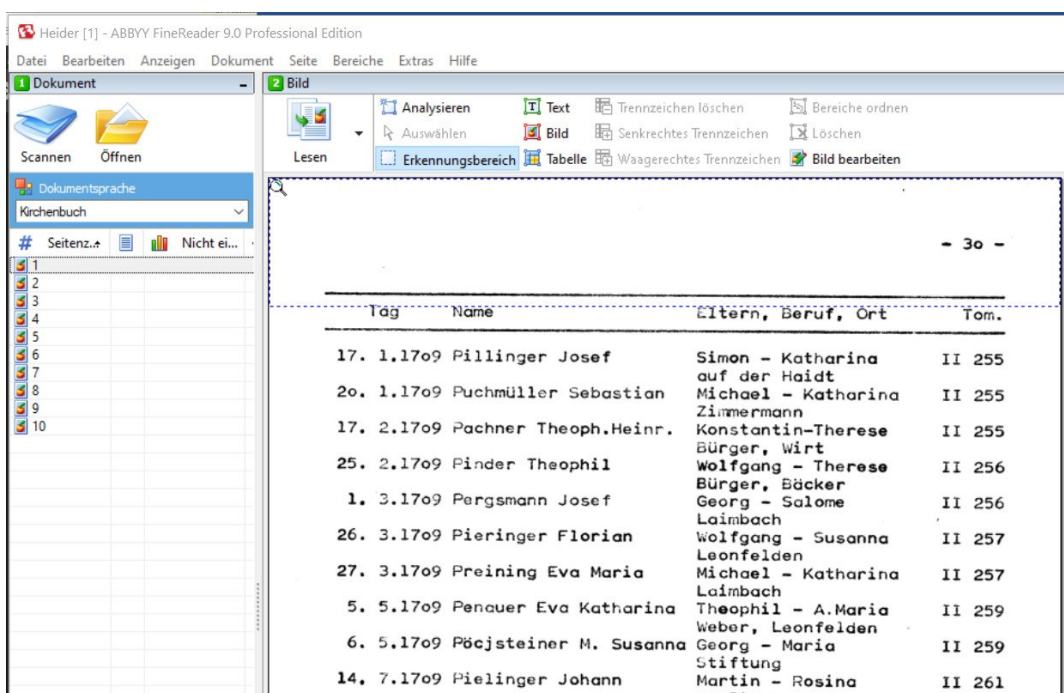


Abb. 3: Öffnen von 10 Registerseiten

Mit der Funktion **Öffnen** wird das gerade aktuelle Verzeichnis (Ordner) mit den darin enthaltenen Scan-Dateien angezeigt. In unserem Fall sollte der Name des Verzeichnisses **Leonfelden_GH_A-G_Scan** sein. Über das Icon mit dem Pfeil nach oben neben dem Suchfeld kann man ein anderes Verzeichnis suchen. Ob man jeweils Einzeldateien (Registerseiten) oder gleich mehrere zur Verarbeitung öffnet, ist eine freie Entscheidung des Bearbeiters. Für mich ist die optimale Menge 10 Seiten. In obiger Ansicht habe ich vom Taufregister A-G von Bad Leonfelden die Seiten 30 bis 39 mit der Funktion **Öffnen** in den FineReader geladen. Mit der Lösch-Taste können einzelne oder alle Seiten wieder entfernt werden. Im Fenster **1 Dokument** des FineReader sieht man die Anzahl der geöffneten Seiten und den Verarbeitungsfortschritt (siehe *Abb. 4: Auswahlrahmen erstellen* auf Seite 13).

3.4 Bereich der Texterkennung

ABBYY FineReader kann auf eine Textstruktur eingehen. So können beispielsweise spaltenorientierte Texte auch spaltenweise bearbeitet werden. Wie wir sehen, ist der Registerseite in 4 Spalten gegliedert. Weil die Spaltengrenzen jedoch nicht immer eingehalten werden, ist eine automatische spaltenweise Texterkennung nicht sinnvoll. Daher wird die ganze Registerseite bzw. der ganze Scan als **ein** Textbereich definiert und verarbeitet. Das genaue Aufteilen der erkannten Texte in die einzelnen Spalten erfolgt erst nachträglich mit einem eigenen Texteditor (siehe *Starten der Formatierung und Prüfung einer Registerseite* auf Seite 21).

Der Erkennungsbereich kann im Fenster **2 Bild** auf 2 Arten definiert werden:

- Klicken auf das grüne Symbol für **T Text** und aufziehen des Texterkennungsrahmens mit der Maus.
- Laden eines vordefinierten Bereiches unter **Bereiche/Bereichsvorlage laden...** Vorher muss ein passender Rahmen mit der Funktion **Bereichsvorlage speichern...** gespeichert worden sein.

Nach Aufziehen bzw. nach Laden des Bereichsrahmens muss dieser in der Regel noch genau justiert werden. Dabei ist darauf zu achten, dass

- alle Texte innerhalb des Bereichsrahmens liegen,
- der Rahmen möglichst genau am linken Rand, also bei der ersten Stelle des Datums, beginnt. Ist dort der Abstand zwischen Rahmen und erstem Zeichen zu groß, fügt FineReader unerwünschte Leerzeichen ein, wobei der FineReader bei einem nach rechts geneigten Bild empfindlicher reagiert als bei einem nach links geneigten. Bei diesem kann am unteren Ende einer Seite die Eintragung um gut eine Stelle weiter rechts beginnen, ohne dass unerwünschte Leerzeichen eingefügt werden. Durch Verzerrungen durch die Schreibmaschine und/oder den Scan-Vorgang können die Scans schief sein. Das Ausrichten von zu schiefen Scans kann mit einem Fotoprogramm erfolgen, darauf wird hier nicht eingegangen.

Die genaue Einstellung des Rahmens ist nur am **linken** Rand wichtig. Rechts, oben und unten können freie Bereiche bleiben. Graue Linien vom Seitenrand der Registerseite oder Abbilder von Fingern, mit denen die Seite beim Scannen gestrafft wird, sollten jedoch möglichst außerhalb des Rahmens liegen, weil sie Störung verursachen können. Normalerweise werden sie aber ignoriert.

Der Rahmen kann als Ganzes durch Gedrückt-halten der Taste **Steuerung** und gleichzeitiges Bewegen der Maus verschoben werden. Die Ränder können aber auch einzeln verschoben werden wie bei einem normalen Windows-Fenster. Vor allem bei einem vordefinierten Auswahlrahmen sollte man nicht vergessen zu kontrollieren, ob unten alle Zeilen innerhalb des grünen Auswahlfeldes liegen. Manche Buchseiten können länger sein als üblich und dann sind die Zeilen möglicherweise nicht mehr innerhalb des Auswahlfeldes und würden bei der Texterkennung verloren gehen.

Der Erkennungsrahmen lässt sich desto genauer justieren, je größer das Bild-Fenster ist. Am unteren Rand des Erkennungsfensters kann man dessen Darstellungsgröße einstellen. Mit der **F6**-Taste nütze ich die ganze Breite des **2 Bild** Fensters. Man muss dann zwar mehr herumschrollen, um die Ränder zu kontrollieren, aber mit dem Mausrad geht das ja sehr flott und der Mehraufwand lohnt sich.

The screenshot shows the ABBYY FineReader 9.0 Professional Edition interface. The main window displays a table of church records. A selection box is drawn around the row for '13.10.1721 Brungraber Leopold'. The table has the following columns: Tag, Name, Eltern, Beruf, Ort, and Tom. The text is as follows:

Tag	Name	Eltern, Beruf, Ort	Tom.
17. 9.1720	Pesenböck Katharina	Simon - Eva Ebretschlag	II 370
7.11.1720	Pamer Elisabeth	Johann - Anna Maria Amessschlag	II 371
7.11.1720	Pachner Leopold	Michael - Katharina Bürger	II 371
27. 2.1721	Frieschl Elisabeth	Michael - Rosina Ebretschlag	II 373
19. 3.1721	Pachner Maria Josefa	Joh.Michael - A.Magdal. Ratsbürger	II 374
29. 3.1721	Pesenböck Anna	Wolfgang - Sophia Amessschlag	II 375
8. 5.1721	Pichler Franz	Josef - Elisabeth Bürger	II 376
27. 5.1721	Pichler Johann	Andreas - Maria Oberstiftung	II 377
7. 9.1721	Pernauer Adam Heinr.&	Joh.Mich. Josef-Kath. Bürger	II 379
9.10.1721	Pogner Martin	Georg - Apollonia Silberhartschlag	II 379
13.10.1721	Brungraber Leopold	Thomas - Salome Bürger	II 379
16.11.1721	Proier Maria	Mathias - Maria Langbruck	II 380
19. 1.1722	Prandstetter A.Maria	Thomas - Susanna Maurer,Dietrichschlag	II 382
21. 1.1722	Piel Anna Maria	Georg - Helene Bürger	II 382

A vertical yellow bar on the right side of the window contains the text: 'Klicken Sie hier, um die Erkennungsergebnisse zu bearbeiten (F8)'.

Abb. 4: Auswahlrahmen erstellen

Manchmal beginnt das Datum um eine Stelle zu weit links. Hier ist es besser, den Rahmen so zu ziehen, dass diese eine Stelle aus dem Rahmen herausfällt. Damit wird vermieden, dass bei allen anderen richtigen Zeilen eine falsche Leerstelle eingefügt wird. Bei der Kontrolle ist so ein Datum um 1 Stelle verschoben und fällt dadurch sofort auf. Durch Einfügen des abgeschnittenen Zeichens wird das Problem rasch behoben. Siehe [Abb. 5: Tag außerhalb des Erkennungsr Rahmens](#) auf der nächsten Seite.

In obigem Beispiel hingegen beginnt das Datum 9.10.1721 durch einen Tippfehler um eine Stelle zu früh. Hier ist nach der OCR-Erkennung ein Leerzeichen einzufügen, um die ganze Zeile um eine Stelle nach rechts zu schieben.

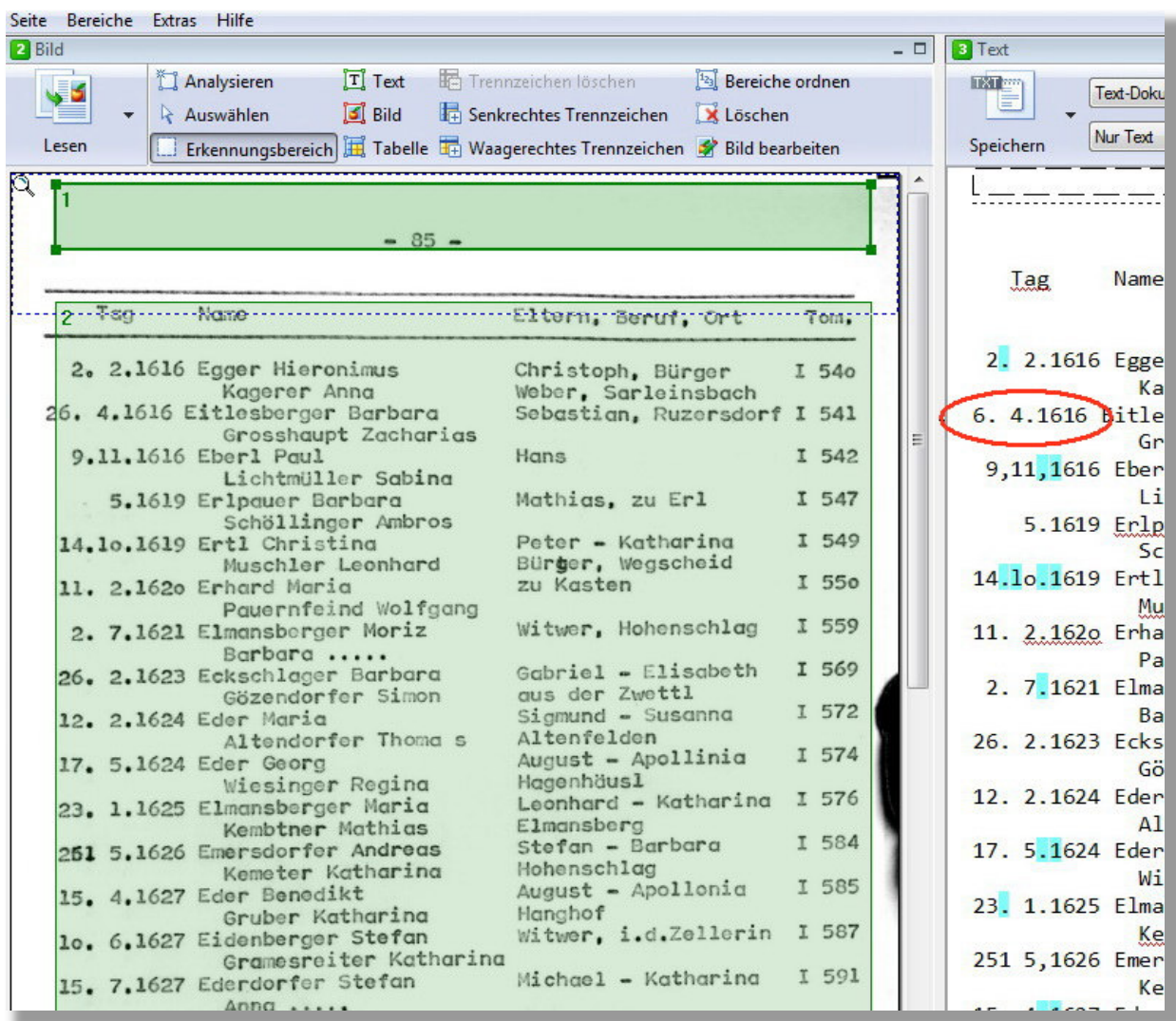


Abb. 5: Tag außerhalb des Erkennungsrahmens

3.5 OCR-Erkennung

Durch Klick auf die Funktion **Lesen** wird die OCR Schrifterkennung für die ausgewählten Seiten gestartet. Im Fenster **1 Dokument** kann man den Fortschritt des OCR-Prozesses mitverfolgen (siehe [Abb. 6: lesen und bearbeiten der Registerseiten](#) auf Seite 15). Je bearbeiteter Seite wird die Gesamtzahl der Zeichen angezeigt. Die unsicheren Zeichen werden in absoluter Zahl und als Prozentwert angezeigt. Ein Prozentsatz von 1 bis 3 % unsichere Zeichen ist ein guter Wert. Nach Ende des OCR-Prozesses werden durch Drücken der **F8**-Taste Seiten mit den umgewandelten Texten im Erkennungsfenster **3 Text** angezeigt. Die unsicheren Zeichen sind blau unterlegt, unbekannte Wörter sind rot unterstrichen. Es sind dann zwar oft viele Zeichen blau unterlegt, aber nur wenige sind tatsächlich falsch und müssen korrigiert werden. Die meisten blau markierten Zeichen sind entweder richtig oder werden durch das nachfolgende Prüfprogramm automatisch korrigiert. Siehe Kapitel [blau markierte Zeichen, die nicht korrigiert werden müssen](#) auf Seite 15.

3.6 Arbeiten mit den umgewandelten Registerseiten

3.6.1 Zoom-Fenster

Für das Bearbeiten des Textes nach der OCR-Erkennung sollte unbedingt das Zoomfenster eingeschaltet sein. Es befindet sich standardmäßig im unteren Teil des Bildschirms und wird über **Anzeigen/Zoomfenster/Zoomfenster anzeigen** aktiviert. Das wird aber bereits beim Einrichten des FineReaders so eingestellt (siehe [OCR-Schrifterkennung mit ABBYY FineReader](#) auf Seite 36).

Sobald im Bild- oder Textfenster ein Zeichen angeklickt wird, wird der Originaltext im Zoomfenster dargestellt. Damit kann der Originaltext mit dem Ergebnis der Umwandlung sofort verglichen werden. Die Vergrößerung der Darstellung lässt sich unabhängig von der Vergrößerung im Fenster **2 Bild** bzw. Fenster **3 Text** einstellen. Handschriftliche Korrekturen, die ja berücksichtigt werden sollen, sind oft nur im **2 Bild** - Fenster gut zu lesen. In solchen Fällen schalte ich mit der **F6**-Taste auf das Bildfenster um und kehre danach mit der **F8**-Taste wieder zum Textfenster zurück.

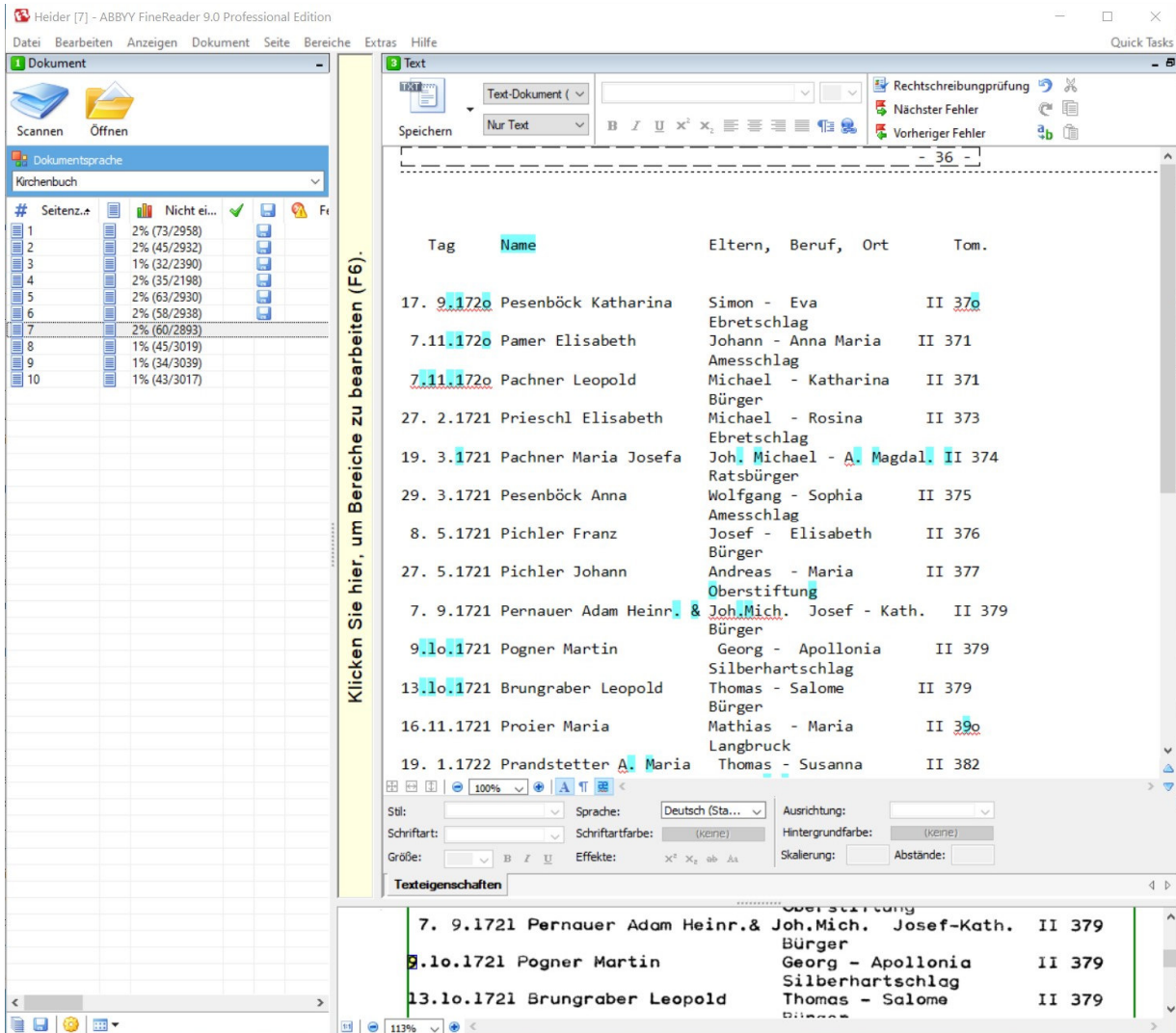


Abb. 6: lesen und bearbeiten der Registerseiten

3.6.2 blau markierte Zeichen, die nicht korrigiert werden müssen

Folgende unsichere (blau markierten) Zeichen benötigen keine Korrektur, weil sie entweder richtig sind oder im nächsten Bearbeitungsschritt durch das Programm automatisch korrigiert werden:

- alle richtig umgewandelten Zeichen, die als unsicher markiert sind
- in der gesamten Registerseite werden folgenden Zeichen vom Formatierungsprogramm (KEDIT) korrigiert:
 - - _ ~ in -
 - ; in :
- in der Dateinummer werden folgenden Zeichen korrigiert:
 - * in -
 - 0 o @ in 0

- I L l in 1
- in der Datumsspalte:
 - Interpunktionszeichen zwischen Tag, Monat und Jahr: werden automatisch auf . gesetzt
 - O o c @ in 0
 - I L l in 1
- im tomus:
 - 1 i l in I
 - e in E
 - x in X
 - Y y v ¥ 7 in V
 - G 0 in C
 - 3 ß in B
- in der pagina:
 - O o @ in 0
 - I l i in 1

3.6.3 Manuelle Prüfung der umgewandelten Texte

Bei der Prüfung der umgewandelten Texte gehe ich spaltenweise vor, d. h. ich konzentriere mich zuerst nur auf die Datumsspalte, dann auf die Namen, danach auf die EBO-Spalte und zuletzt auf die Angaben für tomus und pagina. Diese Vorgehensweise hat sich für mich als die geeignetste herausgestellt, andere mögen ein anderes Vorgehen vorziehen. Beim tomus sollte man prüfen, ob er auch in jeder Zeile angegeben ist. Vor allem bei langen Texten in der EBO-Spalte hat Heider ihn öfters weggelassen, wodurch das Prüfprogramm die pagina, so sie vorhanden ist, als tomus auffasst und ins Schleudern kommt. Bei der manuellen Einfügung gibt es insofern eine kleine Erleichterung, als die Buchstaben I, V, E, X als Kleinbuchstaben eingegeben werden können, weil sie vom Prüfprogramm auf Großbuchstaben umgewandelt werden. Das spart das lästige Betätigen der Umschalttaste auf der Tastatur.

Abhängig von der Schriftqualität der Registerseite gibt es bestimmte Zeichen, die besonders anfällig für Fehlinterpretationen sind bzw. bei denen das OCR-Programm eine falsche Sicherheit vorspielt. Typisch dafür ist die 3, die ich daher immer überprüfe, unabhängig davon, ob FineReader sie blau markiert hat oder nicht. Auch die 5 ist zwar nicht so oft aber doch auch ein Problemzeichen. Beim Text sind m und rn oft verkehrt interpretiert. e, o, c sind weitere Kandidaten, die oft Korrekturen erfordern. Durch die Nachbearbeitung der jpg-Dateien mit einem Fotoprogramm lässt sich das aber wesentlich verbessern.

3.6.4 Prüfung von Namen und Begriffen mit dem Wörterbuch von FineReader

Viele falsche Namen und Begriffe werden erkannt, weil sie im Wörterbuch von FineReader nicht vorkommen. Das Arbeiten mit diesem Wörterbuch und seine Pflege sind daher ganz wichtig und werden in einem eigenen Kapitel beschrieben. Siehe *Benutzerwörterbuch von FineReader* auf Seite 37.

3.6.5 Leerzeichen

Häufig fehlen Leerzeichen zur Trennung von zusammengesetzten Vornamen, z.B. nach einem abgekürzten Vornamen wie A.Maria oder nur um Platz zu sparen (Joh.Michael). Fehlende Leerzeichen bei Vornamen werden bei der späteren Aufteilung der Daten in den Programmen von Dr. Peter Haas richtig erkannt. Es entstehen aber durch zusammengezogene Vornamen ungültige Begriffe, die FineReader rot unterstreicht. Man kann zwar solche zusammengezogene bzw. abgekürzte Namen als eigene Begriffe ins Benutzerwörterbuch speichern, aber ich füge fehlende Leerzeichen immer ein, weil es den Text übersichtlicher macht und die vi-

suelle Kontrolle erleichtert. Dasselbe gilt auch für unterschiedliche Begriffe, die durch einen Beistrich getrennt sind, z.B. Beruf und Ort: Maurer, Dietrichschlag. Auch hier dient ein eingefügtes Leerzeichen der Übersichtlichkeit. Andererseits muss bei den Vornamen der Eltern vor und nach dem Bindestrich ein Leerzeichen stehen, um sie von einem Doppelnamen zu unterscheiden:

Joh.Michael-A.Magdal. → Joh. Michael - A. Magdal.

Genauso muss auch zwischen unterschiedlichen Begriffen ohne trennenden Beistrich, z.B. zwischen Name und tomus, ein Leerzeichen eingefügt werden: A.Magdal.II 374 → A. Magdal. II 374 (siehe Beispiel [Abb. 4: Auswahlrahmen erstellen](#) auf Seite 13, Eintragung vom 19.3.1721)

Keine Korrektur ist erforderlich, wenn ein EBO-Text um **eine** Stelle vor der Startposition der Elternspalte beginnt, sofern vor dieser Stelle ein Leerzeichen steht, also der Text nicht Teil der Namensspalte ist. Da dies öfters vorkommt, wird dieser Fehler vom nachfolgenden Formatierungsprogramm toleriert,

Josef Heider hat manchmal zusätzliche Leerzeichen eingetippt. Sie entstehen aber auch durch die OCR-Umwandlung selbst, weil bei langen Abfolgen von Leerstellen die Texterkennung außer Tritt geraten kann. Das führt bei den letzten beiden Textblöcken oft zu Verschiebungen und die Textblöcke beginnen dann nicht mehr exakt an der vorgesehenen Position. Mehrfache Leerzeichen zwischen Wörtern und führende Leerzeichen werden später automatisch entfernt und brauchen daher nicht gelöscht zu werden.

3.6.6 Trennung von Begriffen (Namen, Beruf, Ort) auf zwei aufeinanderfolgende Zeilen

Ein Begriff kann auf zwei aufeinanderfolgende Zeilen mit einem Abteilungsstrich aufgeteilt sein. Das Programm von Peter entfernt den Abteilungsstrich und fügt dadurch beide Teile zusammen, aber nur, wenn der Teil nach dem Abteilungsstrich mit einem Kleinbuchstaben beginnt. Ist dort ein Großbuchstabe, so handelt es sich nicht um eine Worttrennung, sondern um einen Doppelnamen.

3.6.7 Bearbeitung rückgängig machen

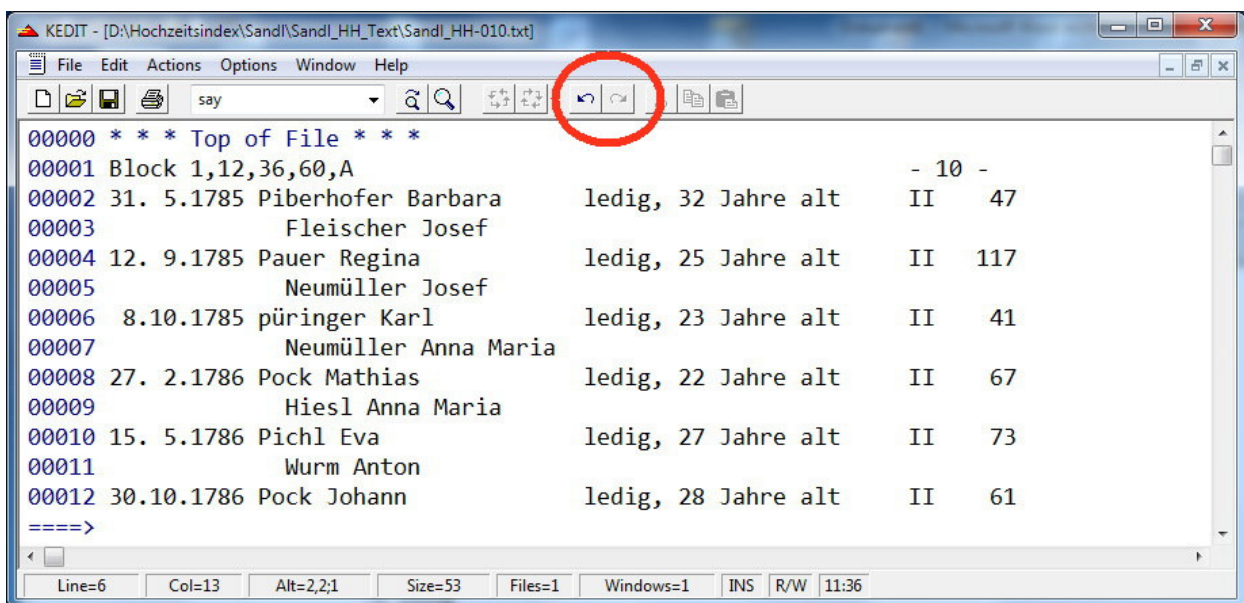


Abb. 7: OCR Textänderungen rückgängig machen

Änderungen des Textes können durch Klick auf die Icons rechts oben rückgängig gemacht bzw. wieder aktiviert werden.

3.6.8 Speichern der bearbeiteten Text-Datei (Registerseite)

Nach der manuellen Korrektur der Textseite ist diese für den nächsten Verarbeitungsschritt zu speichern. Das geschieht mit der Funktion **Speichern**. Wie beim Öffnen muss man auch hier zumindest beim ersten Mal das richtige Verzeichnis aufrufen, bevor man die Datei speichert. In unserem Beispiel ist das das Verzeichnis **Le-onfelden_GH_A-G_Text**. Der Name der Datei muss gleich sein mit dem Namen der Scan-Datei. Um dies zu

erreichen, muss im untersten Auswahlfeld Dateioptionen **Dateien als Quell-Bilddateien benennen** ausgewählt sein. Der Dateityp muss **.txt** sein. Wichtig ist noch, dass das Kästchen **Dokument nach Speichern öffnen** angehakt ist. Alle diese Einstellungen sollte durch das Einrichten des FineReader in den Optionen bereits fix so definiert sein. Der Cursor steht bereits auf dem **Speichern**-Button, sodass ein Mausklick genügt, um die Datei zu speichern und anschließend mit KEDIT zur Formatierung und Datenprüfung aufgerufen wird. Durch einfaches Drücken der Enter-Taste wird dieser Prozess durchgeführt.

3.6.9 Unterbrechung der OCR-Bearbeitung

In *Abb. 6: lesen und bearbeiten der Registerseiten* auf Seite 15 sieht man das Ergebnis bzw. den Fortschritt der OCR-Erkennung der 10 geöffneten Scan-Dateien und welche der Dateien schon gespeichert wurden (Diskettensymbol). Wegen Müdigkeit oder aus anderer Gründe möchte man vielleicht die Arbeit unterbrechen. Das ist jederzeit möglich, weil FineReader in einer internen Datei den aktuellen Zustand zwischenspeichert. Beim nächsten Aufruf des Programmes wird der aktuelle Stand angezeigt und man kann die Arbeit ohne Informationsverlust an der abgebrochenen Stelle fortsetzen.

4 Formatierung und Datenprüfung der Registerseiten mit dem Texteditor KEDIT

4.1 KEDIT

Ich habe schon erwähnt, dass die 4 teilweise unregelmäßig aufgeteilten Spalten der Registerseiten in eine einheitliche Aufteilung gebracht werden müssen und dass die Daten zu prüfen sind. Am Beginn des Projektes mit den Heiraten habe ich das manuell gemacht, aber z.B. die Prüfung des Datums ist so eintönig, dass ich schon nach wenigen Seiten Fehler nicht mehr gesehen habe. Es entstand daher sehr schnell die Idee, für diese Tätigkeiten ein geeignetes Werkzeug einzusetzen. Das fand ich in dem programmierbaren PC-Editor **KEDIT**. Damit kann man sowohl die Daten am PC direkt editieren, als auch mit der sehr umfangreichen Programmiersprache Macros zur Prüfung und Manipulation der Daten erstellen. Die wichtigsten Funktionen, die man zum Manipulieren der Registerseiten am PC braucht, sind in einem eigenen Kapitel beschrieben (*Die wichtigsten Editor Zeilenbefehle* auf Seite 41).

Mit zunehmender Kenntnis der Möglichkeiten der Makrosprache wurden die Prüf- und Formatierungsprogramme immer umfangreicher. Das war auch deshalb notwendig, weil Josef Heider seine Register in oft immer neuen Varianten erfasst hat. Dieser Variantenreichtum hat sich erst im Fortschritt des Projektes gezeigt und erforderte immer wieder Anpassungen der Programme. Diese Anpassungen führten zu einer Qualitätssteigerung der bearbeiteten Textseiten und in weiterer Folge zu einer Beschleunigung im nächsten Projektschritt. Auf die Macro-Sprache gehe hier nicht ein, aber um die von mir entwickelten Macros zu benutzen, müssen sie zuerst auf dem eigenen PC gespeichert und der PC-Editor KEDIT muss auf dem PC installiert werden. Siehe *Installation der KEDIT-Software und der Verarbeitungs-Programme (Macros)* auf Seite 41).

4.2 KEDIT-Profile

```

00033 * =====
00034 *      allgemein gültige Makros starten
00035 * =====
00036 'define f2.kex'          /* toggel Bottom Toolbar */
00037 'define f3.kex'          /* Suchfunktion */
00038 'define f4.kex'          /* Kommentareintragungen suchen */
00039 'define f5.kex'          /* XEDIT Profile ändern */
00040 'define f8.kex'          /* Formalprüfung */
00041 'define f10.kex'         /* zurückblättern */
00042 'define f11.kex'         /* vorwärtsblättern */
00043 'define Enter.kex'      /* Formalprüfung */
00044 if MatrikelTyp = 'HH' then do
00045   'define f6.kex'         /* Elternblock rechts verschieben */
00046   'define f7.kex'         /* pagina nach rechts verschieben nur bei Matrikeltyp=Heirat */
00047   'define f9.kex'         /* Gegeneitragung suchen (nur bei Matrikeltyp=Heirat) */
00048 end
00049 if MatrikelTyp = 'SH' then 'set toolbar on both'
00050 /* In den folgenden Variablen werden die Startpositionen der 4 Blöcke (Spalten) gesetzt */
00051 /* Bei Beginn der Verarbeitung einer neuen Pfarre sind diese Werte im Profile zu überprüfen, */
00052 /* weil diese Positionen in den Heider-Registerseiten variieren können. */
00053 /* tomBlock schwankt durch unkontrolliertes Einfügen von Leerstellen durch die OCR-Umwandlung stark */
00054 /* häufig verschoben wird und damit zu Fehlern führt. */
00055 *trace +r
00056 'editv set DatumBlock 1'      /* Beginn des Datums */
00057 'editv set NameBlock 12'     /* Beginn des Probanden */
00058 'editv set ElternBlock 36'   /* diese Position schwankt von Buch zu Buch und muss angepasst werden */
00059 'editv set tomBlock 58'      /* muss auch an die Spalteneinteilung der Registerseite angepasst werden */
00060 'editv set ElternBlockMin 40' /* neue Position des ElternBlock, darf nicht < ElternBlock sein! */
00061 '*editv set tomBlockMin 70'   /* neue Position des tomBlock */
00062 if MatrikelTyp = 'SH' then 'editv set tomBlockMin 70' /* neue Position des tomBlock bei Sterbe 75 */
00063 else 'editv set tomBlockMin 70' /* sonst 70 */
00064 'editv set ParkPos 100'      /* auf diese Position werden der Eltern- und tomBlock temporär zwischengeparkt */
00065 *
00066 /* Die Reihenfolge von Tomus und Pagina ist in manchen Büchern verkehrt und wird hier definiert
00067 'editv set TomSequence B'     /* Reihenfolge: Pag Tom bzw. Pag/Tom */
00068 'editv set TomSequence C'     /* Pag unterhalb von Tom oder im Elternteil */
00069 'editv set TomSequence A'     /* Reihenfolge: Tom Pag (Normalfall) */
00070 *call trace off
00071 /* In der folgenden Tabelle sind alle Tom-Nummern definiert, die in der zu bearbeitenden
00072 /* Pfarre vorkommen. Ihre Reihenfolge definiert die Sortierfolge, unabhängig von dem tom-Numer.
00073 /* Die römischen Nummern bzw. Buchstaben sind durch eine Leerstelle zu trennen.
00074 AllTom='I'
00075 'editv put AllTom'
00076 end
00077 * * * End of File * * *
=====

```

Abb. 8: Ausschnitt aus dem KEDIT-Profile

Für die Formatierung und für die Prüfung der Daten sind einige Informationen notwendig, die in Form von Parametern im sog. **KEDIT-Profile** gespeichert sind. Einige davon sind spezifisch für die zu bearbeitenden

Registerseite und müssen vor Beginn des Prozesses an den Aufbau der Registerseiten angepasst werden. Dieses Profile wird beim Aufruf von KEDIT durchlaufen und speichert dabei die Informationen in Variablen, die dann von den Verarbeitungs-Macros benützt werden.

4.2.1 Startposition der 4 Spalten (Blöcke)

In obiger Abbildung des Profiles wird in den Zeilen 00055 bis 00058 definiert, auf welchen Positionen die 4 Spalten (Blöcke) beginnen. Beim ElternBlock und tomBlock kann es, abhängig von der Einstellung der Schreibmaschine, mit der Josef Heider die Registerseiten geschrieben hat, leichte Verschiebungen geben. Diese Werte sind daher gegebenenfalls anzupassen.

4.2.2 Reihenfolge von tomus und pagina

In den Heider-Büchern gibt es drei Varianten, wie tomus und pagina angeordnet sein können. In der Variablen TomSequence wird die aktuelle Reihenfolge definiert. Im Profile sieht dies wie folgt aus:

```
'editv set TomSequence A'      /* Reihenfolge: Tom-Pag (Normalfall) */
'editv set TomSequence B'      /* Reihenfolge: Pag-Tom */
'editv set TomSequence C'      /* Pagina unterhalb von Tom */
```

In der Version C gibt es noch Unterformen, die vom Programm analysiert werden müssen, um die pagina herauszufiltern. Diese verschiedenen Formen sind allerdings bis Ende 2024 nur bei den Heiraten, aber nicht bei den Taufen vorgekommen.

4.2.3 Liste der tomus-Nummern

Ebenfalls anzupassen ist die Liste der tomus-Nummern in AllTom. Das sind die Nummern der Matrikenbücher einer Kategorie in chronologisch aufsteigender Reihenfolge. Sie sind oft auf einer der Anfangsseiten eines Registerbuches angegeben, auf alle Fälle aber im Verzeichnis der Kirchenbücher *Heider-Register: Verzeichnis der erfassten Matrikenbücher nach Pfarren*, das Josef Heider angelegt hat und auf den Netzseiten der Familia Austria aufgerufen werden kann ¹.

```
AllTom=' I II III IV'
```

Die Nummern sind ein- oder mehrstellige Zeichen bzw. Begriffe, die durch eine Leerstelle getrennt sind. Meistens sind es römische Nummern, aber nicht immer. In jedem Fall sind sie so einzugeben, wie sie Josef Heider in seiner Aufstellung je Kategorie definiert hat, denn so hat er sie auf den Registerseiten unter Tom. angeführt.

Die Profile-Datei befindet sich im selben Verzeichnis wie die KEDIT-Macros und kann von dort direkt aufgerufen und modifiziert werden. Der Aufruf ist aber auch bei einer mit KEDIT geöffneten Register-Datei mit der **F5**-Taste möglich. Nach Veränderung der Profile-Datei nicht vergessen, die Registerseite zu schließen und dann neu zu öffnen, damit der Profile neu durchlaufen wird und die Änderungen wirksam werden!

4.3 Aufruf von Funktionen für die Tauf- und Sterberegister über Funktionstasten

Es gibt 7 Funktionen, die über Funktionstasten aufgerufen werden.

F2 Ein-/Ausschalten bottom toolbar (der top toolbar ist immer sichtbar), siehe *.Arbeiten mit Textblöcken* auf Seite 42.

F3 Mit dieser Funktion kann man ein ganzes Registerbuch oder Teile davon **Seite für Seite** nach einem Begriff durchsuchen (kann auch aus mehreren Wörtern bestehen), um ihn zu überprüfen oder auch, um ihn zu ändern. Die Suche ist case-sensitiv, der Suchbegriff muss also mit Groß- und Kleinbuchstaben korrekt im Suchfenster eingegeben werden. Mit dem v-Zeichen in der Suchzeile kann man aus der drop down Liste auch früher verwendete Suchbegriffe auswählen. Mit den Pfeiltasten links und rechts bzw. oben und unten

¹ <https://www.familia-austria.at/index.php/component/search/?searchword=heider&searchphrase=all&Itemid=110> (Link erstellt Ende 2024)

auf der Tastatur wird nach dem nächsten oder vorhergehenden Begriff gesucht. Mit der Eingabetaste sucht man in der aktuellen Richtung weiter. Die Suche wird durch Klick mit der Maus beendet.

F4 Nach Kommentaren suchen.

Damit kann man rasch prüfen, ob es noch Registereintragungen mit einem Kommentar gibt, die einer weiteren Nachprüfung, z.B. über matricula, bedürfen. Mit den Pfeiltasten links und rechts kann man die Suchrichtung bestimmen.

F5 XEDIT-Profile aufrufen.

Damit können Variable im Profile auf die Werte gesetzt werden, wie sie für die Textprüfung benötigt werden (siehe *KEDIT-Profile* auf Seite 19).

F6, F7 nur bei Heiratsregistern

F8, Enter Formatierung und Prüfung einer Registerseite

F10 rückwärtsblättern im aktuellen Registerbuch

F11 vorwärtsblättern im aktuellen Registerbuch

4.4 Starten der Formatierung und Prüfung einer Registerseite

Unmittelbar nach der OCR-Umwandlung und Speichern der Registerseite wird diese mit KEDIT zur Bearbeitung aufgerufen. Voraussetzung dafür ist, dass Dateien mit dem Dateityp **txt** mit der Anwendung **KEDIT** geöffnet werden. Das ist im Betriebssystem mit der Funktion **Standard-Apps nach Dateityp auswählen** festzulegen. Eine gespeicherte Registerseite kann aber auch später jederzeit geöffnet und bearbeitet werden. Beim ersten Aufruf hat die Seite noch das Format, wie sie Josef Heider mit der Schreibmaschine erstellt hat inkl. der Korrekturen, die nach der OCR-Umwandlung vorgenommen wurden (siehe *Abb. 6: lesen und bearbeiten der Registerseiten* auf Seite 15). Durch Drücken der Eingabe- oder F8-Taste startet der Formatierungsprozess.

Im Unterschied zu den Taufregistern müssen bei den Sterberegistern vor dem Starten der Formatierung der tomus-Block und die Altersangaben so eingerichtet werden, dass sie von den Informationen im EBO-Block völlig getrennt sind. Außerdem muss dem Programm mitgeteilt werden, an welcher Position sich der tomus befindet, indem der Cursor dort positioniert wird. Siehe die genaue Beschreibung im Kapitel *tomus-Block und Altersangabe beim Sterberegister* auf Seite 25.

4.5 Lesen der vorhergehenden Seite

Für manche Prüfungen ist die letzte Registereintragung der vorhergehenden Seite mit dem gleichen Registerbuchstaben maßgeblich. Am Beginn des Prüfprozesses wird daher diese Eintragung gesucht. Wird keine vorhergehende Seite gefunden, weil z.B. die aktuelle Seite die erste des Registerbuches ist oder hat die vorhergehende Registerseite einen anderen Index-Buchstaben, wird durch eine Nachricht darauf hingewiesen.

4.6 Seitennummer

Vor und nach der Seitennummer muss ein Minus-Zeichen sein. Die Seitennummer ist zwingend notwendig und muss mit der Seitennummer im Dateinamen übereinstimmen. Bei fehlender Seitennummer oder bei Nichtübereinstimmung bricht das Programm mit einer Nachricht ab, ohne dass Daten verändert werden.

Abweichungen zwischen Seitennummer des Registerbuches und der Dateinummer in den Scan- und Text-Dateien entstehen durch fehlerhafter Seitennummerierung im Heider-Register, z.B. durch Doppelvergabe derselben Seitennummer oder Auslassung von Seitennummern. Die Dateinummer muss an die falsch nummerierte Registerseite angepasst werden. Damit man auch im Nachhinein weiß, welche Seitennummer Heider ursprünglich vergeben hat, ist diese in der Textdatei nach der berichtigten Seitennummer zwischen runden Klammern anzugeben.

Mehr zum Thema Seitennummer findest du im Dokument *Heider-Projekt technische Dokumentation* im Kapitel *Fehler in der Nummerierung der Heider-Registerseiten und ihre Behandlung im Dateinamen*. Siehe auch Innenseite des Titelblattes.



Abb. 9: Die Seite 171B wurde eingeschoben, weil Heider die Nr. 171 zweimal vergeben hat

4.7 Überschriftzeile

Die Überschrift über den Spalten beginnt mit dem Wort **Tag**. Diese Zeile wird automatisch erkannt und gelöscht. Die Zeile kann aber auch zur Gänze fehlen. Ein manueller Eingriff ist nur erforderlich, wenn die Überschriftzeile nicht mit dem Schlüsselwort **Tag** beginnt. Die Texte danach sind irrelevant.

4.8 Leerzeilen

Am Beginn einer Registerseite sind bis zum Beginn der ersten Eintragung ein oder mehrere Leerzeilen. Diese werden durch das Programm automatisch gelöscht. Auch Leerzeilen zwischen den Eintragungen werden dort gelöscht und benötigen daher keinen manuellen Eingriff.

4.9 Eintragung

Eine Eintragung besteht in der Regel aus 2 Zeilen, von denen die zweite oft leer ist. Bei umfangreicheren Eintragungen können es auch 3 oder mehr Zeilen sein. Die erste Zeile endet mit dem **tomus**, in den Folgezeilen kann der Text noch in den **tom**-Block reichen. Ob das wirklich ein relevanter Text ist, muss geprüft werden.

4.9.1 Name

Beginnt der Name auf einer späteren Position als 12, erfolgt eine Fehlermeldung. Diese wird im Programm bereits generiert, bevor es führende und mehrfache Leerstellen im Textteil entfernt. D.h., dass zwar eine Fehlermeldung erscheint, aber bei einer neuerlichen Formatierung der Name richtig positioniert wird bzw. bereits ist. Daher: Einfach nochmals die Eingabetaste drücken und die Fehlermeldung verschwindet. (Ich habe das bisher aus purer Faulheit nicht geändert.)

Der Anfangsbuchstabe muss ein Großbuchstabe sein und mit dem Indexbuchstaben der Seite übereinstimmen. Vor dem Namen kann ein **v.**, **von**, **de**, oder eine **(** stehen. Der Anfangsbuchstabe steht dann entsprechend weiter rechts. Bei der Prüfung auf Übereinstimmung ist das spezielle Alphabet im Heider-Index zu beachten, siehe im Dokument *Heider-Register und Digitalisierungsprojekt Einführung* im Kapitel *Alphabetischer Index*. Siehe auch Innenseite des Titelblattes.

Ein Register-fremder Anfangsbuchstabe wird durch eine Fehlermeldung und durch gelbe Einfärbung der Zeile angezeigt. Nichtübereinstimmung kann entweder durch einen Fehler in der OCR-Umwandlung entstehen, die dann zu korrigieren ist, oder Josef Heider hat einen Namen erwischt, der zu einem anderen Index gehört. Das passiert z.B. manchmal, wenn er einen Namen mit **St** unter **S** oder **Sch** etc. einordnet oder die Eintragung schlicht verreicht hat. In diesem Fall tragen wir einen Kommentar ein und lassen die Eintragung unverändert stehen.

Eine besondere Situation entsteht, wenn der Anfangsbuchstabe in der ersten Eintragung der Registerseite nicht mit dem Index übereinstimmt. In diesem Fall werden alle anderen, richtigen Namen mit der Fehlermeldung versehen, sie würden mit einem falschen Anfangsbuchstaben beginnen und die Eintragungen werden gelb markiert. Damit gehen echte Fehler in den vielen gelb markierten Zeilen unter. Daher übertippe ich den Anfangsbuchstaben in der ersten Eintragung mit dem aktuellen Indexbuchstaben, prüfe die Registerseite wie üblich und mache die ‚falsche Korrektur‘ des Anfangsbuchstaben erst vor dem Speichern der Seite wieder rückgängig.

4.9.2 Vertauschte Vor- und Familiennamen

Es kommt vor, dass Josef Heider in der Namensspalte die Reihenfolge Familienname – Vorname vertauscht hat. Hier ist die Reihenfolge richtig zu stellen. Dasselbe gilt auch, wenn er die Namens- und die EBO-Spalte vertauscht hat.

4.9.3 Sonderfälle beim Namen

4.9.3.1 *Alternative Schreibweise des Familiennamens*

- 32 -

Tag	Name	Eltern, Beruf, Ort	Tom.
21.10.1597	Bell Eva	Georg, Eckersberg	AI 3
9. 2.1698	Pausch Johann	Wolfgang, Inwohner	AI 4
15. 3.1598	Peiter Georg	Mathias, Altenfelden	AI 4
4. 7.1598	Perntaller Simon	Adam, Weber, zu Edt	AI 6
8. 8.1698	Pausch Maaria	Stefan, Azesberg	AI 7
11.10.1598	Pernecker Barbara	Paul, von Lanfeldt	AI 8
3. 5.1599	Pocksrucker Florian	Georg, Kazenbach	AI 14
2. 9.1599	Plaumauer Maria	Mathias, Altenfelden	AI 16
14. 9.1599	B(w)iblinger Mathias	Urban, Altenfelden	AI 17
17. 6.1600	Pauer Thomas	Hauns, Neundling	AI 25

Abb. 10: alternative Schreibweise des Familiennamens

Der Familienname kann in dieser Form mit einer Suchfunktion nur sehr schwer gefunden werden. Daher ist die 2. Namensvariante in voller Länge, aber in Klammern, nach der ersten Variante anzuführen:

14. 9.1599 Biblinger (Wiblinger) Mathias Urban, Altenfelden AI 17

4.9.3.2 *Vulgo-Namen, 2 Familiennamen*

Manchmal sind beim Kind der Familienname und ein Vulgo-Name angegeben:

4.11.1629 vulgo Schramm Leopold A
gen. Puderer Thomas

20. 5.1630 vulgo Schwarzbart Hans - Ursula A
gen. Wallmann Johann

Josef Heider hat in diesen Fällen das Kind, so wie bei unehelichen Kindern, unter beiden Namen in seinem Register verzeichnet. Damit findet man es praktischerweise unter sowohl als auch.

Es gibt auch Fälle, in denen beim Kind 2 Familiennamen ohne vulgo angeführt werden:

5. 5.1633 Schwarzbart oder Wallmann Tobias Hans - Ursula A

Vermutlich handelt es sich auch hier um einen vulgo-Namen, es wurde aber nicht explizit angegeben. Diese Fälle werden bei der Aufteilung der Daten behandelt bzw. sind durch den Forscher oder die Forscherin zu klären.

4.9.3.3 *Sohn, Tochter, Kind*

Vor allem bei Totgeburten oder kurz nach der Geburt verstorbenen Kindern finden sich die Bezeichnungen **Sohn**, **Tochter** oder **Kind**. Mit Sohn und Tochter ist zumindest das Geschlecht bestimmt. Den Begriff **Kind** bitte durch 5 Punkte (.....) ersetzen.

4.10 Eltern, Beruf, Ort (EBO)

In der Elternzeile kann es bereits zu kleinen Verschiebungen des Textes kommen, sodass der Text schon eine Stelle vor oder nach der definierten Position des Elterntextes stehen kann, sofern davor ein Leerzeichen ist (sonst wäre es ja Teil des Namens). Da eine einstellige Linksverschiebung des Elterntextes öfters vorkommt, wird das vom Programm toleriert. Beginnt der Elterntext weiter rechts, werden die führenden Leerstellen eliminiert, genauso wie mehrfache Leerstellen zwischen den Textteilen. Nach erfolgter Formatierung beginnt der Elterntexte genau an der vorgesehenen Position. Abgekürzte Namen kann man in eindeutigen Fällen ergänzen, z.B. Schwertbg. in Schwertberg, um sie bei der DB-Suche leichter auffindbar zu machen.

4.11 Zahlenangaben in der EBO-Spalte

Manchmal kommen in der EBO-Spalte Zahlen vor, z.B. ein Sterbedatum. Die evt. falschen Zeichen l und o werden vom Formatierungsprogramm nicht automatisch erkannt und korrigiert. Diese Korrektur muss händisch erfolgen. Auch eingebettete Leerzeichen sind händisch zu entfernen.

4.12 tomus-Block

tomus, die Nummer des Buches, ist zwingend. Wenn man das Fehlen des tomus übersieht, werden die nächsten Zeichen (pagina), so vorhanden, als tomus interpretiert. Das ergibt dann eine ungültige tomus-Nummer und führt zu einer entsprechenden Fehlermeldung.

Wenn zur Klärung von Unsicherheiten durch uns eine Registereintragung im Originalbuch aufgesucht wird, dort eine Seitennummer steht, die Heider nicht in seine Registereintragung übernommen hat, wird diese fallweise in der Registerseite nachgetragen. Die steht dann allein auf weiter Flur⁴ und ist sozusagen ein Service für die Benutzer.

Die normale Form ist *tomus pagina* (sofern die Seitennummer angegeben ist). Manchmal ist die Reihenfolge umgedreht: *pagina/tomus* (dann oft, aber nicht immer, durch einen Schrägstrich getrennt). Die Reihenfolge kann mitten in einer Buchseite wechseln. Hier sind oft manuelle Eingriffe notwendig. Das war bei den Heiraten oft der Fall, bei denen es noch komplexere Formen gibt, aber kaum bei den Taufen.

4.12.1 tomus-Block beim Taufregister

Zunächst muss das Programm wissen, wo die Tom.-Spalte beginnt. Bei den Taufen wird im ersten Formatierungsprozess dazu die im KEDIT-Profil definierte Zahl in der Variablen `tomBlock` verwendet. Diese Zahl definiert also die Position der tomus-Spalte. Das Programm geht zunächst 3 Stellen weiter nach links und sucht von dort nach rechts das erste Zeichen, das kein Leerzeichen ist. Der dort beginnende Text wird als tomus aufgefasst. Durch diese Vorgehensweise wird auch ein tomus erkannt, der weiter rechts steht, als in der Variablen `tomBlock` definiert ist.

Bei diesem Verfahren muss der Cursor allerdings vor Position 54 oder nach 63 stehen, also außerhalb des Bereiches von tomus-Block. Ist auf einer Seite der tomus zu weit links, weil z.B. FineReader durch sehr viele Leerstellen in der EBO-Spalte irritiert wurde, kann man alternativ die tomus-Position mit dem Cursor definieren, indem man ihn an den Beginn des am weitesten linksstehenden tomus auf der Seite stellt. Das Suchen nach der tomus-Nummer beginnt dann von dieser Position aus. Man kann das leicht ausprobieren, da man Änderungen jederzeit rückgängig machen kann (siehe *Rückgängig machen von Änderungen* auf Seite 32).

Steht der tomus in einer Eintragung zu weit rechts, ist das so lange kein Problem, als kein Einzelzeichen, z.B. ein Bindestrich aus der EBO-Spalte, dazwischenkommt. In diesem Fall muss man händisch eingreifen.

Da in der pagina, die nach dem tomus folgt, Kleinbuchstaben vorkommen können, wird **c nicht in 0** umgewandelt. Ist tomus absteigend oder ungültig oder die pagina innerhalb desselben tomus absteigend, wird eine Fehlermeldung ausgegeben und die Zeile gelb markiert. Bei Wiederholung des Formatierungslaufes ist die Position des Tom-Blockes bereits bekannt, da in der Kopfzeile gespeichert (siehe *Blockzeile in der Text-Datei* weiter unten). Bei neuerlichem Aufruf des Programmes ist daher die Position des Cursors ohne Bedeutung.

Ein absteigender tomus ist in der Regel ein Tippfehler, vereinzelt auftretende absteigende pagina sind meist durch Verreihung der Eintragungen entstanden oder ebenfalls durch Tippfehler. Gehäuft hin- und herspringende pagina treten als Folge der unseligen Reform der Kirchenbücher von 1784 auf, wonach die Eintragungen nach Ortschaften zu erfolgen habe. Diese Bestimmung hatte nur Nachteile und wurde bald wieder zurückgenommen. Ihre Anwendung führt dazu, dass im Heider-Register die pagina ständig hin- und herspringt. Jeder Rücksprung in der pagina führt zu einer Fehlermeldung und das können dann auf einer Seite sehr viele sein. Durch Eingabe des Textes **P** bzw. **p** (pagina) in der Befehlszeile (gelb eingefärbte Zeile am unteren Fensterrand nach dem ==>) werden diese Fehlernachrichten unterdrückt. Andere Fehler wie z.B. absteigendes Datum werden aber weiterhin angezeigt. Die Unterdrückung der Fehlermeldungen wird in der Blockzeile der Registerseite vermerkt und gilt damit für alle weiteren Aufrufe zur Formatierung dieser Seite.

4.12.2 tomus-Block und Altersangabe beim Sterberegister

Für den tomus-Block in den Sterberegistern trifft das vorhin bei den Taufregistern Gesagte ebenfalls zu. Die Besonderheit in den Sterberegistern gegenüber den Heiraten und den Taufen besteht aber in der optionalen Angabe des Alters, die vor dem tomus, aber am Ende des Elternblocks steht. Sie besteht aus einer Zahlenangabe und dem Buchstaben T für Tag, W für Wochen und M für Monate. Ist kein Buchstabe angegeben, sind mit der Zahlenangabe Jahre gemeint. Diese Altersangabe muss nun vor der Erstformatierung, also wenn die Registerseite nach der OCR-Verarbeitung mit KEDIT zum ersten Mal aufgerufen wird, von der EBO-Information getrennt werden. Da das Programm nicht zwischen EBO-Information und der Altersangabe unterscheiden kann, müssen der tomus-Block und die Altersangaben manuell weiter nach rechts verschoben werden. Dazu ist die Funktion *Arbeiten mit Textblöcken* wie auf Seite 42 beschrieben, sehr hilfreich.

Der für das Verschieben von Blöcken benötigte bottom toolbar wird beim Aufrufen eines Sterberegisters mit KEDIT bereits automatisch eingeblendet, kann aber auch mit der F2-Taste ein- und ausgeblendet werden. Die Verschiebung nach rechts muss nun so weit erfolgen, dass der Abstand zwischen dem Ende der EBO-Information und dem tomus mindestens 6 Stellen beträgt. Ist bei einer Eintragung kein Alter angegeben, kann die EBO-Information bis kurz vor dem tomus stehen und wird dann mitverschoben. Bei dem folgenden Beispiel ist das in den Zeilen 10, 32, 44 und 48 der Fall. Das muss manuell rückgängig gemacht werden.



Abb. 11: Textblock mit Alter und tom-Block nach rechts verschoben

Altersangabe und tomus dürfen sich nicht überlappen. Solche Fälle müssen ebenfalls händisch korrigiert werden. Außerdem hat Josef Heider in diesem Beispiel die Altersangabe aus Platzgründen in die Folgezeilen 33 und 49 geschrieben. Das lässt sich durch Selektieren und durch Ziehen mit gedrückter Maustaste an die gewünschte Position in der Datumszeile korrigieren.

Damit können wir die erste Formatierung starten. Wir stellen den Cursor auf die am weitesten linksstehende tomus-Position. In unserem Fall ist das Position 62 – z.B. in der Zeile 24. Ich habe sie in der folgenden Darstellung durch die rote Linie markiert. Die Altersangabe ist innerhalb der blauen und der roten Linie isoliert.

00010	11. 9.1675	Autengruber Martha	Totengräber, Wächter	II	248
00011					
00012	4. 6.1678	Amprosin Maria	Praitenstein	II	254
00013					
00014	6. 6.1682	Almesberger Kind	Florian, Schindlau	II	261
00015					
00016	23. 3.1684	Autengruber Andreas	Rudolfing	11	265
00017					
00018	2.12.1688	Almesberger Adam	Aigen	II	274
00019					
00020	14. 9.1715	Almesberger Kaspar	Schindlau	6o	III 496
00021					
00022	22. 6.1716	Almesberger M. Magdalena	Schindlau	5	III 498
00023					
00024	9. 6.1718	Almesberger Maria	Wurmbrand	6	III 5o5
00025					
00026	6.12.1719	Asanger Johann	Knecht zu Aigen	3o	III 5o9
00027					
00028	5. 5.1723	Aigner Lukas	Landgerichtsdieners zu Aigen	6o	IV 589
00029					
00030	24. 1.1725	Almesmüller Philipp	wurmbrand	7o	IV 595
00031					
00032	26. 8.1726	Almesberger Magdalena	Hans Georg, Schindlau	15M	IV 6o3
00033	9. 7.1787	Amgner Hans Georg	ledig, Aigen	28	IV 6o9
00034					
00035	6. 2.1729	Asanger Maria	Michael, ledig zu Unterneudorf	11W	IV 613
00036					
00037	7. 4.1729	Almesberger Gertrude	Hans Georg	7	IV 615
00038					
00039	22.12.1729	Asanger Margarete	Holzhäuseln	6o	IV 619
00040					
00041	1o. 3.173o	Aigner Eva Maria	Landgerichtsdieners zu Aigen	71	IV 621
00042					
00043	17. 9.173o	Aigner M. Katharina	Kaspar, Langerichts- diener zu Aigen	7T	IV 624
00044					
00045	9. 7.1731	Andexlinger Maria	Rudolfing	51	IV 627
00046					
00047	29. 9.1731	Aigner M. Magdalena	Kaspar, Landgerichts- diener zu Aigen	1oW	IV 628
00048					

Abb. 12: EBO-Information und Altersangabe entflechtet

Nach Drücken die Eingabetaste werden die Blöcke neu positioniert, die Seite sauber formatierte und am Beginn der Datei eine Blockzeile erzeugt. Wir sehen: Alles, was unmittelbar oder auch weiter rechts von der Cursor-Position steht, wird als tomus und pagina, was links davon steht, wird als die Altersangabe interpretiert. Da jetzt das Programm zwischen Altersangabe, tomus und pagina unterscheiden kann, können wir Fehler der OCR-Umwandlung gezielt korrigieren. Die Buchstaben beim Alter müssen wir noch kontrollieren. Ein T kann als 1 und ein M als W bzw. umgekehrt umgewandelt worden sein und das ist nur durch manuelle Kontrolle erkennbar.

Line	Date	Name	Location	Page
00000		* * * Top of File * * *		
00001	Block 1,12,40,70,A,	- 2 -		
00002	18.10.1674	Aichinger Stefan	Aigen	II 247
00003	20. 8.1674	Autengruber Michael	Rudolfing	II 247
00004	13. 5.1675	Autengruber Anna		II 248
00005	11. 9.1675	Autengruber Martha	Totengräber, Wächter	II 248
00006	4. 6.1678	Amprosin Maria	Praitenstein	II 254
00007	6. 6.1682	Almesberger Kind	Florian, Schindlau	II 261
00008	23. 3.1684	Autengruber Andreas	Rudolfing	II 265
00009	2.12.1688	Almesberger Adam	Aigen	II 274
00010	14. 9.1715	Almesberger Kaspar	Schindlau	60 III 496
00011	22. 6.1716	Almesberger M. Magdalena	Schindlau	5 III 498
00012	9. 6.1718	Almesberger Maria	Wurmbrand	6 III 505
00013	6.12.1719	Asanger Johann	Knecht zu Aigen	30 III 509
00014	5. 5.1723	Aigner Lukas	Landgerichtsdieners zu Aigen	60 IV 589
00015				
00016	24. 1.1725	Almesmüller Philipp	Wurmbrand	70 IV 595
00017	26. 8.1726	Almesberger Magdalena	Hans Georg, Schindlau	15M IV 603
00018	9. 7.1787	Amgner Hans Beorg	ledig, Aigen	28 IV 609
00019	6. 2.1729	Asanger Maria	Michael, ledig	11W IV 613
00020			zu Unterneudorf	
00021	7. 4.1729	Almesberger Gertrude	Hans Georg	7 IV 615
00022	22.12.1729	Asanger Margarete	Holzhäuseln	60 IV 619
00023	10. 3.1730	Aigner Eva Maria	Landgerichtsdieners zu Aigen	71 IV 621
00024				
00025	17. 9.1730	Aigner M. Katharina	Kaspar, Langerichts- diener zu Aigen	7T IV 624
00026				
00027	9. 7.1731	Andexlinger Maria	Rudolfing	51 IV 627
00028	29. 9.1731	Aigner M. Magdalena	Kaspar, Landgerichts- diener zu Aigen	10W IV 628
00029				
00030	27.12.1731	Aigner Anna Maria	Landgerichtsdieners zu Aigen	24 IV 629
00031				
00032	7. 2.1732	Asanger Thomas	Johann, Zaglau	7W IV 630
00033	28. 3.1732	Almesberger Rosina	Hans Georg, zu Schindlau	4 IV 630
00034				
00035		* * * End of File * * *		

Abb. 13: formatierte Seite des Sterberegisters mit Altersangaben

Das Programm fügt zwischen EBO-Block und der Altersangabe das Trennzeichen | ein. Das ist eine Hilfe beim Auseinanderhalten vom EBO-Text und der Altersangabe, vor allem, wenn diese fehlt. Was rechts vom Trennzeichen und vor dem tomus steht, ist eine Altersangabe, was links davon steht, gehört zum EBO-Text. Nun kann der EBO-Text beliebig manipuliert werden. Dabei kann sich das Trennzeichen nach links oder rechts verschieben, darf aber nicht gelöscht oder überschrieben werden. Fehlt es bei neuerlichen Formatierung, erfolgt eine Fehlermeldung. Man muss dann das fehlende Trennzeichen händisch einfügen und die Formatierung neu durchführen. Die formatierte Registerseite (-datei) wird mit den Trennzeichen an die nächste Verarbeitungsstufe weitergegeben, obwohl es dort vielleicht nicht benötigt wird. Die Angabe für tomBlock in der Blockzeile (ersten Zeile der formatierten Registerseite) betrifft auch hier die Position des tomus-Blockes (und nicht die der Altersangabe). Siehe [Blockzeile in der Text-Datei](#) auf Seite 29.

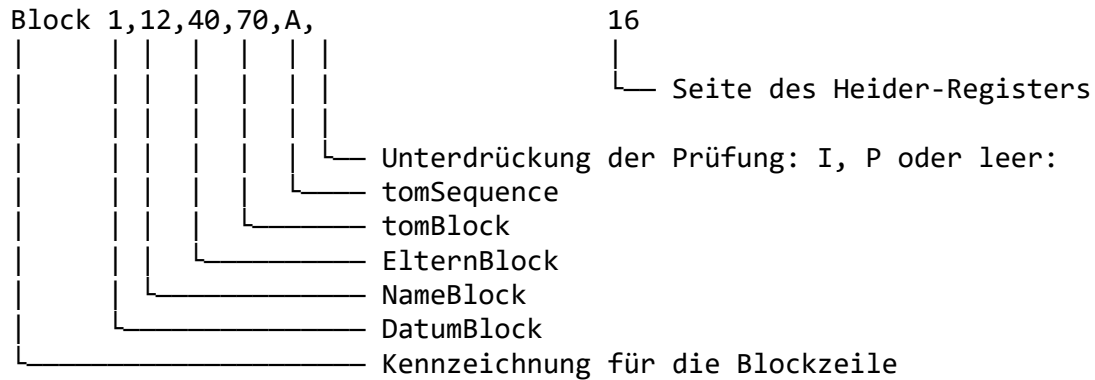
4.13 falsche Zeilen

Manchmal sind Zeilen im Heider-Buch mit dem Wort falsch gekennzeichnet, weil beim Schreiben der Eintragung ein Fehler erkannt wurde. Solche Zeilen sind zur Gänze zu löschen. Im Kommentarfeld der richtigen Zeile ist ein Hinweis einzutragen, z.B. vorhergehende Zeile falsch, gelöscht.

4.14 Manuelle Korrekturen in den Heider-Registern durch Benutzer

Häufig findet man auf den maschinengeschriebenen Seiten der Heider-Bücher händische Eintragungen. Das sind Korrekturen von Benutzern, die Fehler gefunden haben. Wir übernehmen diese Korrekturen nur dann, wenn wir ihre Richtigkeit durch Vergleich mit der Originaleintragung in matricula geprüft haben.

4.15 Blockzeile in der Text-Datei



Nach dem Speichern einer mit dem FineReader umgewandelten Registerseite hat diese noch das Format, wie sie Josef Heider mit der Schreibmaschine erstellt hat inkl. der Korrekturen, die nach der OCR-Umwandlung vorgenommen wurden. Bei der ersten erfolgreichen Formatierung und Datenprüfung mit den KEDIT-Macros wird in der ersten Zeile ein Informationsblock erzeugt mit dem Schlüsselwort **Block** am Zeilenbeginn. Die Informationen in dieser Zeile werden beim wiederholten Formatieren mit KEDIT benützt und auch bei der weiteren Bearbeitung durch Peters Programm verwendet. Die Zahlen definieren die Startposition der verschiedenen Blöcke, der Buchstabe am Ende definiert die Anordnung von tomus pagina. Bei fertigen Seiten ist dieser immer A, weil durch die Bearbeitung mit KEDIT die Normalform tomus pagina hergestellt wird (es gibt also auch bei tomus/pagina Varianten). Dahinter kann noch der Buchstabe für die Fehlerunterdrückung folgen.

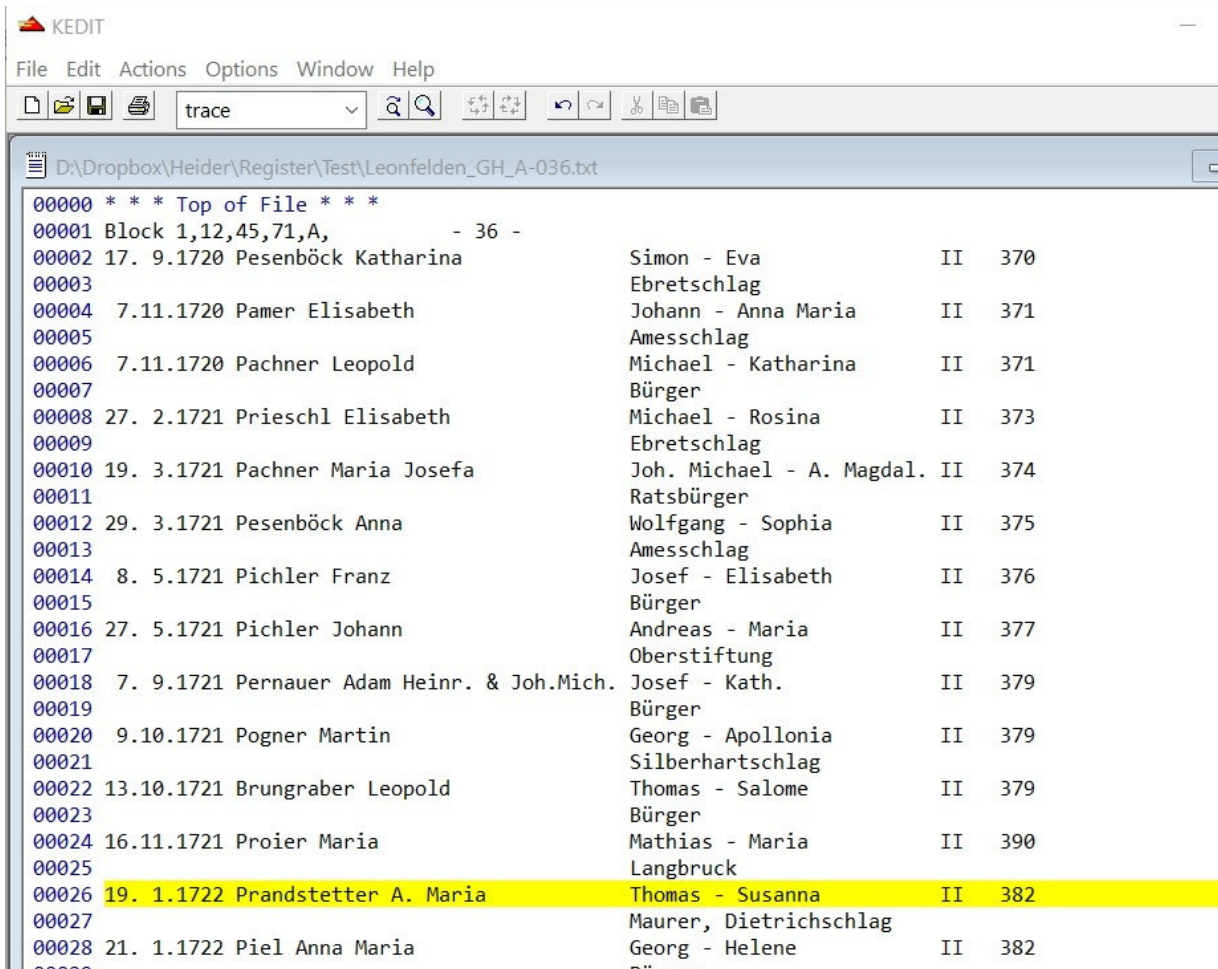


Abb. 14: formatierte Registerseite mit eingefügter Blockzeile

Dem aufmerksamen Leser fällt vielleicht auf, dass die Startpositionen von EBO und von **Tom.** nicht den Minimalpositionen von 40 und 70 entsprechen, wie wir sie im Profile gesehen haben. Ich gestehe, dass ich bei diesem Beispiel gemogelt habe, indem ich beim 2. Zwilling **Joh.Mich.** am 7.8.1721 **kein** Leerzeichen eingefügt habe. Außerdem ist die Zeile **00026** gelb hervorgehoben. Die zugehörige Fehlermeldung habe ich schon entfernt (man hätte sonst die Block-Angaben nicht gesehen). Sie hat darauf hingewiesen, dass die pagina absteigend ist. Der Grund ist der Tippfehler in der vorhergehenden Zeile, wo die pagina 390 mit einer 8 übertippt wurde, FineReader die Zahl aber trotzdem mit 390 übersetzt hat. Dieser Fall war schnell erklärt, bei den Zwillingen muss ich etwas weiter ausholen.

4.16 Zwillinge und mehrfache Vornamen

Es gilt zwar die Regel, dass Begriffe in der 3. Spalte zum EBO-Block gehören – mit einer Ausnahme: Vornamen, die auf Grund ihrer Anzahl und/oder Länge in den 3. Block ragen, gehören zum Namen-Block des Kindes. Das passiert oft bei Zwillingen, die Josef Heider mit einem **&** verbindet (manchmal fälschlicher Weise mit **+**, mit dem aber Heider ein Sterbedatum markiert). Alle Vornamen, die mit einem **&** verbunden sind, platziert KEDIT in der Namensspalte, wodurch sie manchmal länger wird. Es verschiebt dann die EBO- Spalte so weit nach rechts, dass die Namen Platz haben. Bei diesem Prozess darf aber der letzte Vorname nach einem **&** kein doppelter sein, weil dann nicht klar ist, ob ein Teil des letzten Vornamens nicht schon zum Vater gehört. Aus diesem Grund habe ich bei **Joh.Mich.** kein Leerzeichen eingefügt und damit KEDIT vorgeschwindelt, es handle sich nur um **einen** Namen. Man darf dann aber nach der Formatierung der Seite nicht vergessen, die Leerstelle nachträglich einzufügen, sobald die Vornamen in der Namensspalte sind. Dadurch wird dann die Namensspalte noch einmal um eine Stelle breiter, aber auch das erfolgt automatisch bei neuerlichem Formatieren.

Die Verlängerung der Namensspalte muss einheitlich für alle Eintragungen einer Seite gemacht werden. Das Programm benötigt daher 2 Durchläufe, um die längste Namensspalte (und auch die längste EBO-Spalte) zu

ermitteln. Dabei kann es allerdings zu einer sehr großen, unschönen und unpraktischen Breite der Spalten kommen. Durch eine manuelle Verschiebung des Vornamens bzw. der Vornamen in die nächste Zeile (siehe Zeile 18, können auch mehrere Zeilen sein) lässt sich das wieder rückgängig machen. In der nächsten Abbildung sieht man, dass das Formatierungsprogramm durch die Verschiebung die EBO- und die tomus-Spalte auf die Minimumpositionen 40 und 70 zurückgegangen sind. Die Leerstelle zwischen Joh.Mich. ist dann noch einzufügen. Eine ähnliche Situation tritt auf, wenn es nicht um Zwillinge (oder Drillinge), sondern um viele und lange Vornamen eines Einzelkindes geht, wie wir sie manchmal bei Adligen (oder Mächtigen-Adligen) finden. Hier ist dann immer ein manueller Eingriff notwendig. In jedem Fall immer optisch kontrollieren.

```

00000 * * * Top of File * * *
00001 Block 1,12,40,70,A,          - 36 -
00002 17. 9.1720 Pesenböck Katharina      Simon - Eva          II  370
00003                                     Ebretschlag
00004 7.11.1720 Pamer Elisabeth           Johann - Anna Maria II  371
00005                                     Ameschlag
00006 7.11.1720 Pachner Leopold          Michael - Katharina  II  371
00007                                     Bürger
00008 27. 2.1721 Prieschl Elisabeth       Michael - Rosina     II  373
00009                                     Ebretschlag
00010 19. 3.1721 Pachner Maria Josefa      Joh. Michael - A. Magdal. II  374
00011                                     Ratsbürger
00012 29. 3.1721 Pesenböck Anna           Wolfgang - Sophia    II  375
00013                                     Ameschlag
00014 8. 5.1721 Pichler Franz              Josef - Elisabeth   II  376
00015                                     Bürger
00016 27. 5.1721 Pichler Johann           Andreas - Maria      II  377
00017                                     Oberstiftung
00018 7. 9.1721 Pernauer Adam Heinr. &    Josef - Kath.       II  379
00019                                     Joh.Mich.           Bürger
00020 9.10.1721 Pogner Martin              Georg - Apollonia    II  379
00021                                     Silberhartschlag

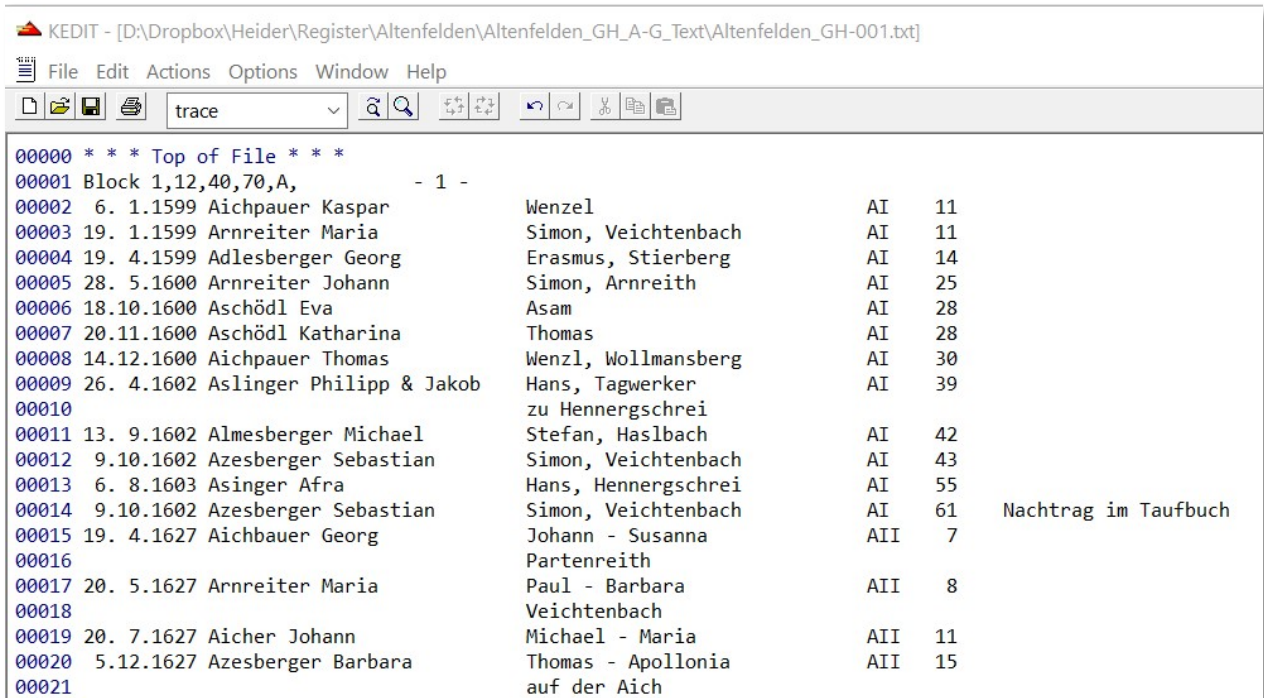
```

Abb. 15: Vornamen in die Folgezeile verschoben

4.17 Fehlermeldungen, Kontrolle

Wir hatten gerade einen Fall mit einem Fehler in einer Zeile. Da es sich hier um einen simplen Tippfehler handelte, wird der stillschweigend ausgebessert. Auch bei eindeutig verschriebenen Vornamen halte ich das so. Es gibt aber Situationen, bei denen erst durch Nachschau im Matrikelbuch eindeutig geklärt werden kann, ob tatsächlich ein Fehler vorliegt. So kann z.B. ein absteigendes Datum ein Tippfehler, oder der Eintrag von Josef Heider beim Einsortieren verreicht worden oder ein Nachtrag im Matrikelbuch sein. Das Nachschauen in matricula kostet aber viel Zeit und stört den Arbeitsfluss. Dzt. schreibe ich nur einen Kommentar in der Kommentarspalte, denn das Ergebnis meiner Arbeit wird von Kontrollern bzw. Kontrolleserinnen (dzt. Marianne Reisinger) kontrolliert. Damit entdeckt sie Fehler, die mir durch die Lappen gegangen sind und prüft Zweifelsfälle. Um das Nachschauen in matricula zu beschleunigen, sortiert sie einen Block von Eintragungen so um, dass sie nach Matrikenbüchern geordnet sind und sie nicht bei jedem Buchstabenwechsel von vorne mit dem Suchen beginnen muss.

In Zeile **00014** sehen wir einen Kommentar nach der **Tom.**-Spalte. Das Prüfprogramm hat ein absteigendes Datum erkannt und gemeldet. Zunächst würde man vermuten, dass das Jahr **1603** oder größer heißen sollte und der Fehler würde nach Korrektur verschwinden. Ob die Vermutung **1603** zutrifft, lässt sich allerdings nur durch Nachschau in der Originalseite mit Sicherheit sagen. Tatsächlich ist nach Aufruf der Seite über matricula zu sehen, dass im Kirchenbuch eine nachträgliche Taufeintragung für das Jahr **1602** vorgenommen wurde. Bei der Kontrolle diese Seite wurde daher ein Kommentar hinzugefügt, um der nachfolgenden Datenaufteilung die Korrektheit der Eintragung mitzuteilen.



```

00000 * * * Top of File * * *
00001 Block 1,12,40,70,A, - 1 -
00002 6. 1.1599 Aichpauer Kaspar Wenzel AI 11
00003 19. 1.1599 Arnreiter Maria Simon, Veichtenbach AI 11
00004 19. 4.1599 Adlesberger Georg Erasmus, Stierberg AI 14
00005 28. 5.1600 Arnreiter Johann Simon, Arnreith AI 25
00006 18.10.1600 Aschödl Eva Asam AI 28
00007 20.11.1600 Aschödl Katharina Thomas AI 28
00008 14.12.1600 Aichpauer Thomas Wenzl, Wollmansberg AI 30
00009 26. 4.1602 Aslinger Philipp & Jakob Hans, Tagwerker AI 39
00010 zu Hennergschrei
00011 13. 9.1602 Almesberger Michael Stefan, Haslbach AI 42
00012 9.10.1602 Azesberger Sebastian Simon, Veichtenbach AI 43
00013 6. 8.1603 Asinger Afra Hans, Hennergschrei AI 55
00014 9.10.1602 Azesberger Sebastian Simon, Veichtenbach AI 61 Nachtrag im Taufbuch
00015 19. 4.1627 Aichbauer Georg Johann - Susanna AII 7
00016 Partenreith
00017 20. 5.1627 Arnreiter Maria Paul - Barbara AII 8
00018 Veichtenbach
00019 20. 7.1627 Aicher Johann Michael - Maria AII 11
00020 5.12.1627 Azesberger Barbara Thomas - Apollonia AII 15
00021 auf der Aich

```

Abb. 16: Meldung im Kommentarfeld

4.18 Nachtrag

Auf der letzten Seite eines Buchstabens steht oft ein Nachtrag, gekennzeichnet durch den Text **Nachtrag**: Die danach folgenden Registereintragungen werden auf gleiche Weise, wie oben beschrieben, geprüft. Datums- und tomus-Prüfungen beginnen also wieder von vorne.

4.19 Rückgängig machen von Änderungen

Änderungen, die manuell oder über ein Programm (Macros) gemacht wurden, können durch Klick auf das linke Pfeilsymbol wieder rückgängig gemacht werden. Z.B. bei fehlendem tomus, den man übersehen hat, kommt die Formatierung etwas durcheinander. Es ist dann oft schneller, die Formatierung rückgängig zu machen, den fehlenden tomus nachzutragen und die Formatierung zu wiederholen.

Mit jedem Klick auf das Symbol werden Änderungen in zeitlicher Reihenfolge rückgängig gemacht. Das ist sehr praktisch, weil man damit Änderungen von mehreren Durchgängen Stufe für Stufe zurücknehmen kann. Ist man eine Stufe zu weit gegangen, kann man das mit Klick auf das rechte Pfeilsymbol wieder korrigieren.

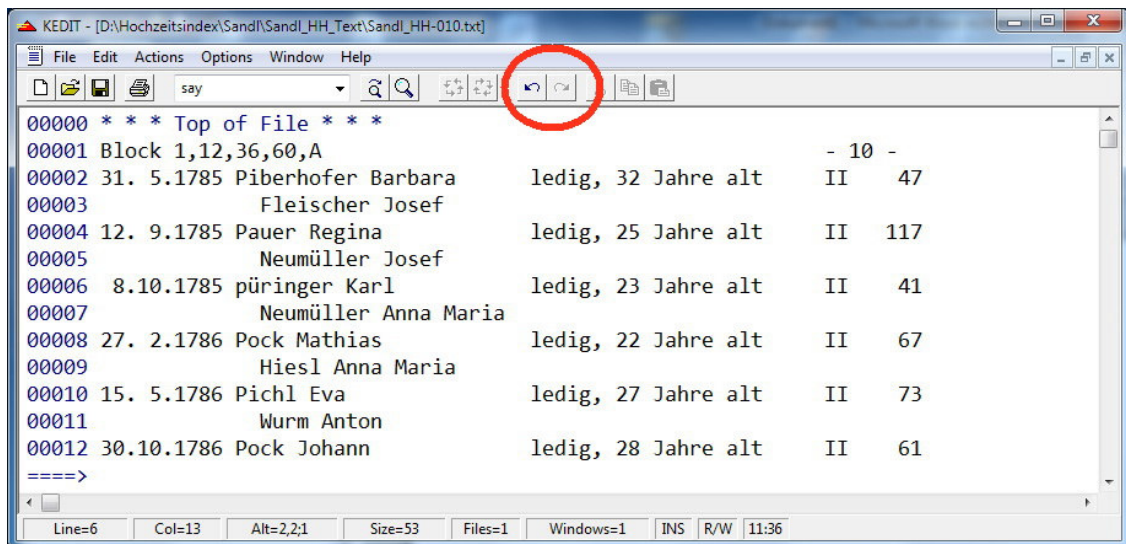


Abb. 17: Rückgängig machen von Änderungen in KEDIT

4.20 Anhang

Am Ende eines Register-Buches gibt es meist eine nicht nummerierte Buchseite mit dem Text **Anhang** und danach ein oder mehrere Seiten mit Eintragungen, die in den Matrikelbüchern fehlerhaft, unleserlich oder unvollständig sind und die Heider nicht richtig zuordnen konnte. Nach der Prüfung sind fast alle Zeilen als fehlerhaft markiert, weil hier kaum bzw. gar keine Familiennamen, sondern nur Vornamen aufgelistet sind und die Anfangsbuchstaben nicht mit dem Indexbuchstaben übereinstimmen. Diese Prüfung und die Fehlermeldungen sind im Grunde unsinnig und daher zu ignorieren. Mit der Eingabe **i** oder **l** (initial) in der Befehlszeile wird die Prüfung des Anfangsbuchstabens unterdrückt und damit die Fehlerliste übersichtlicher. Andere Fehler wie z.B. absteigendes Datum werden aber weiterhin angezeigt. Die Unterdrückung der Fehlermeldungen wird in der Blockzeile der Registerseite vermerkt und gilt damit für alle weiteren Aufrufe zur Formatierung bzw. Prüfung dieser Seite.

5 OCR-Schrifterkennung mit ABBYY FineReader

5.1 Die FineReader-Optionen zum Einrichten und Anpassen

ABBYY FineReader muss an die speziellen Bedürfnisse dieses Projektes angepasst werden. Dies geschieht vor allem mit der Funktion **Extras/Optionen...** Nach Installation des FineReader sind die nachfolgenden Optionen über die Reiter in der obersten Zeile zu wählen.

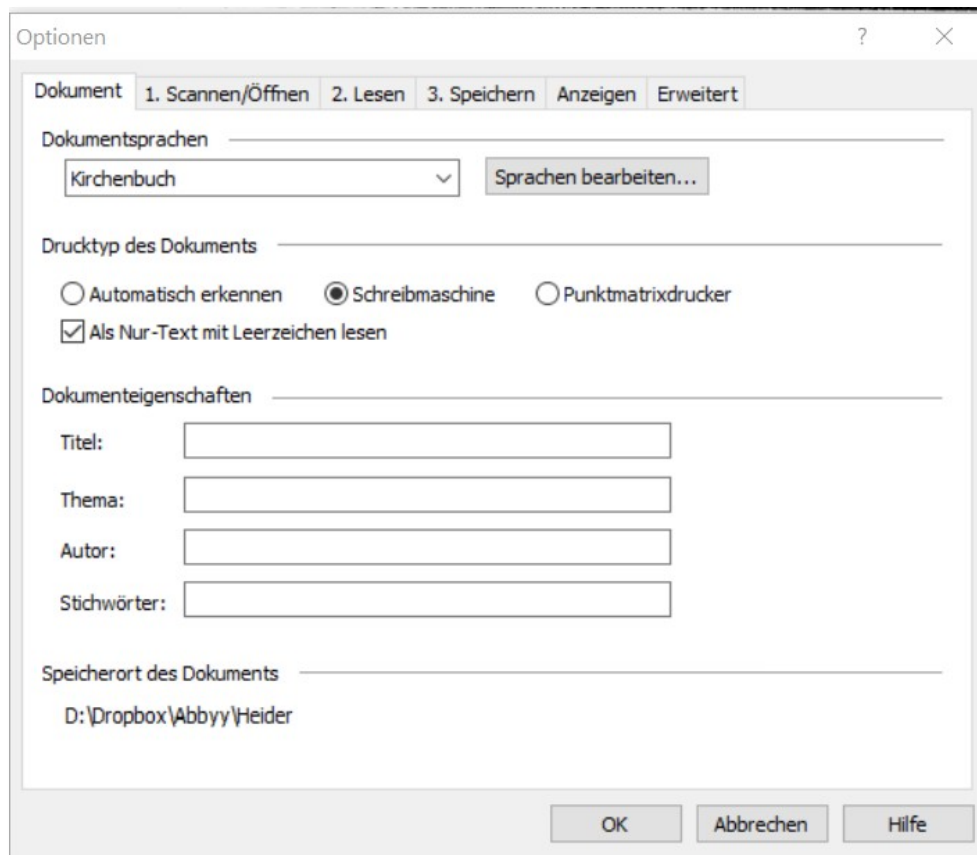


Abb. 18: FineReader-Option Dokument

5.1.1 Dokument

Zu **Dokumentsprachen** gibt es ein eigenes Kapitel (siehe *Benutzerwörterbuch* auf Seite 37). Beim **Drucktyp** bewirkt das Häkchen bei der Auswahl **Als Nur-Text mit Leerzeichen** eine möglichst positionsgenaue Textdarstellung ohne Tabulatoren. Voraussetzung dafür ist die Verwendung einer monospace Schrift für die Darstellung des übersetzten Textes, siehe *Anzeigen: Schriftart für Text* auf Seite 36. Unter **Speicherort des Dokumentes** ist nicht das Verzeichnis der zu übertragenden Dateien gemeint, sondern ein internes Verzeichnis, in dem FineReader unter anderem den Bearbeitungsstand der aktuellen Seiten speichert.

5.1.2 Scannen/Öffnen

Im Reiter **1. Scannen/Öffnen** kann aus drei Möglichkeiten ausgewählt werden, wann der Prozess der Texterkennung beginnen soll. Hier ist die Variante **Erfasste Seitenbilder nicht automatisch lesen...** zu wählen. Andernfalls beginnt der Prozess bereits unmittelbar nach dem Laden der Scans, also noch bevor der Erkennungsbereich manuell definiert werden kann.

Die Option **Bildneigung korrigieren** sollte nicht ausgewählt werden. FineReader richtet die Bilder waagrecht aus. Da durch Verzerrungen der Seite die Zeilen manchmal schief sind, sind die Zeilen zwar waagrecht, aber der linke Rand wird dabei schief. Entscheidend ist jedoch, dass die Zeilen senkrecht untereinanderstehen. Schiefe Seiten sollten vor dem Öffnen mit einem Fotoprogramm geradegerichtet werden.

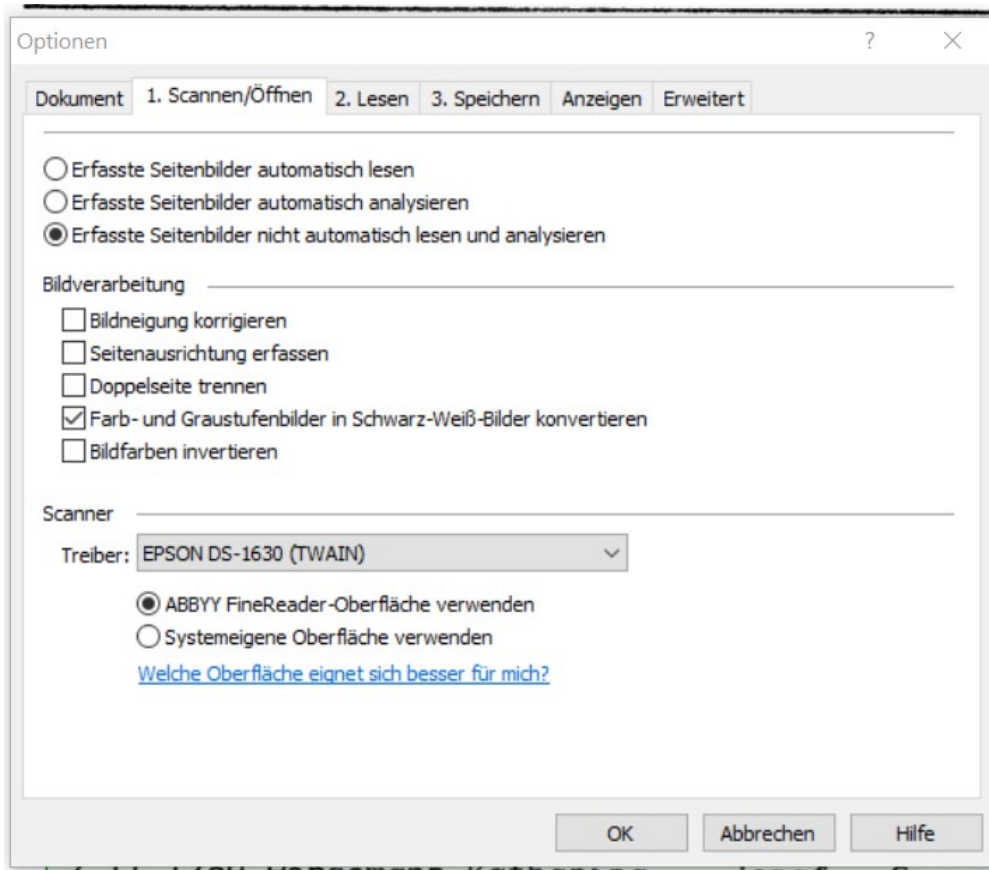


Abb. 19: FineReader-Option Scannen/Öffnen

5.1.3 Lesen: Gründlich, Benutzermuster

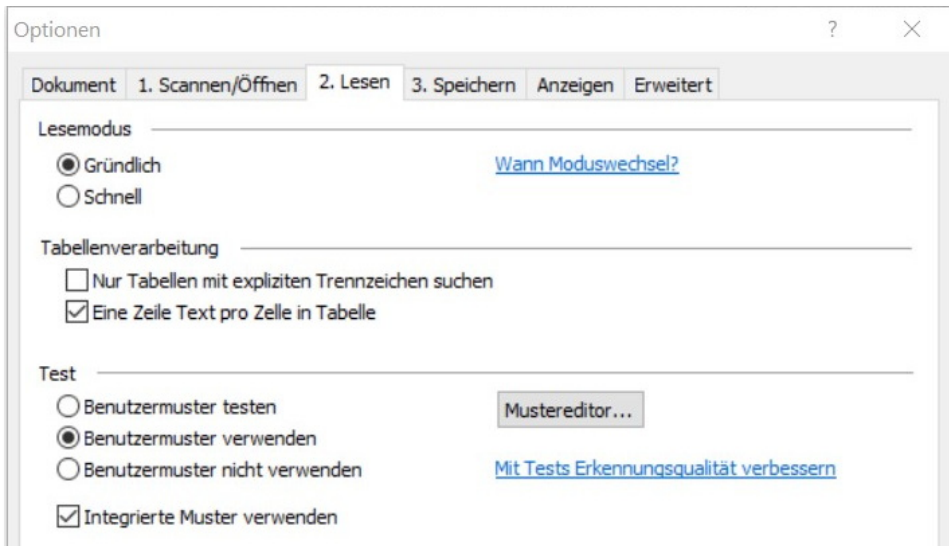


Abb. 20: FineReader-Option Lesen

5.1.4 Speichern: TXT

Der übersetzte Text kann in verschiedenen Formaten abgespeichert werden (RTF, XLS, HTML, PDF, ...) Wir verwenden beim Reiter **3. Speichern** das Format **TXT**.

Kopf- und Fußzeile beibehalten ist auszuwählen, weil FineReader die Seitennummer zu einer Kopfzeile macht, die mit dem Text mitgespeichert werden muss.

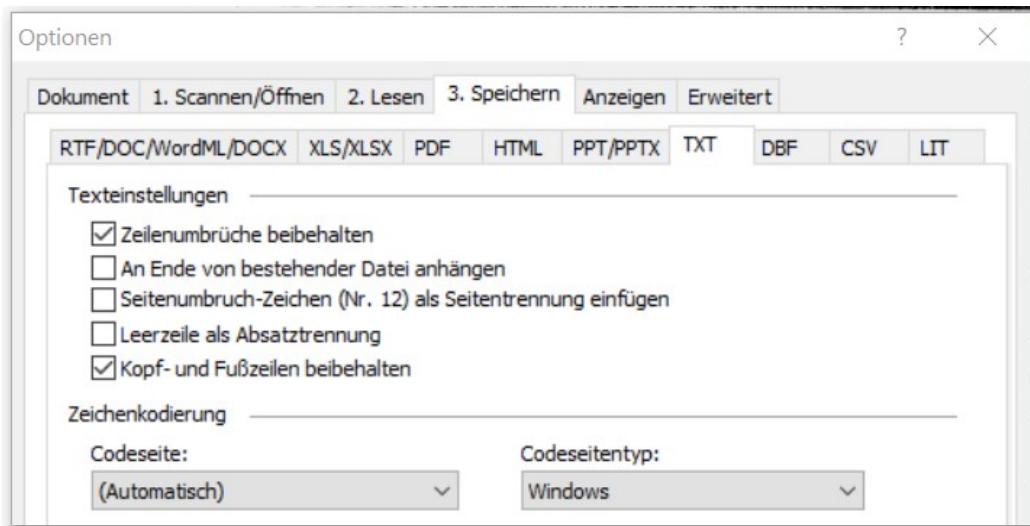


Abb. 21: FineReader-Option Speichern

5.1.5 Anzeigen: Schriftart für Text

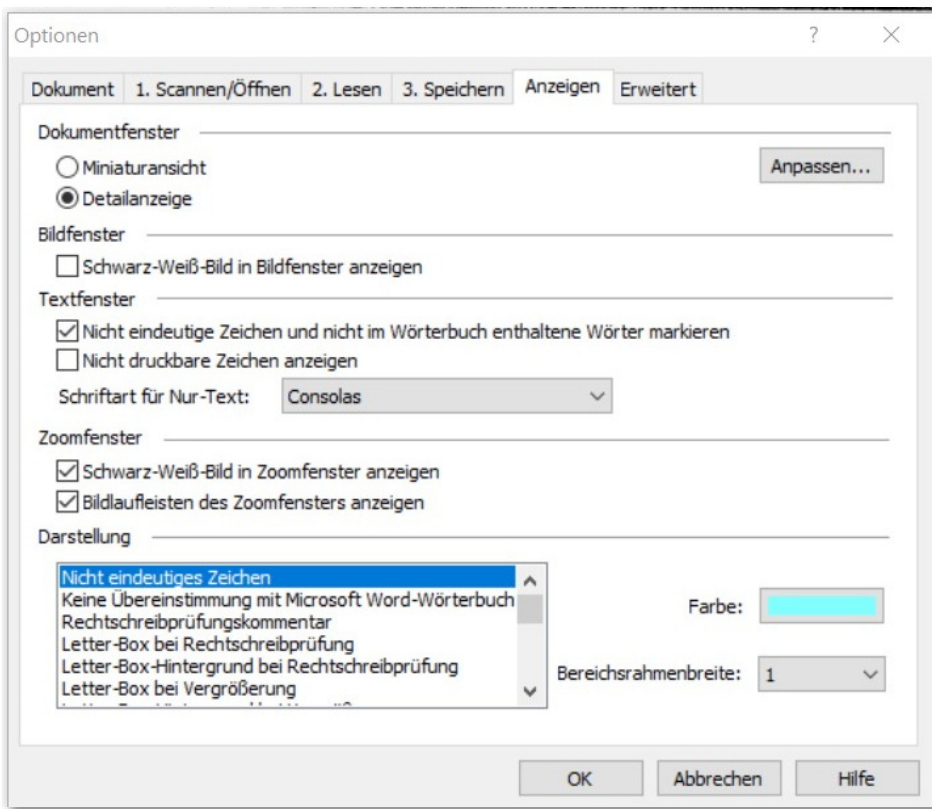


Abb. 22: FineReader-Option Anzeigen

Die von Heider verwendeten Schreibmaschinen sind vom Typ monospace, (jedes Zeichen hat gleiche Buchstabenbreite). Im Reiter **Anzeigen** muss daher für die Textanzeige ebenfalls eine monospace-Schriftart gewählt werden, weil sonst die fixe Spalteneinteilung optisch verloren geht. Ich habe im Reiter **Anzeigen** die Schriftart **Consolas** ausgewählt.

5.1.6 Erweitert

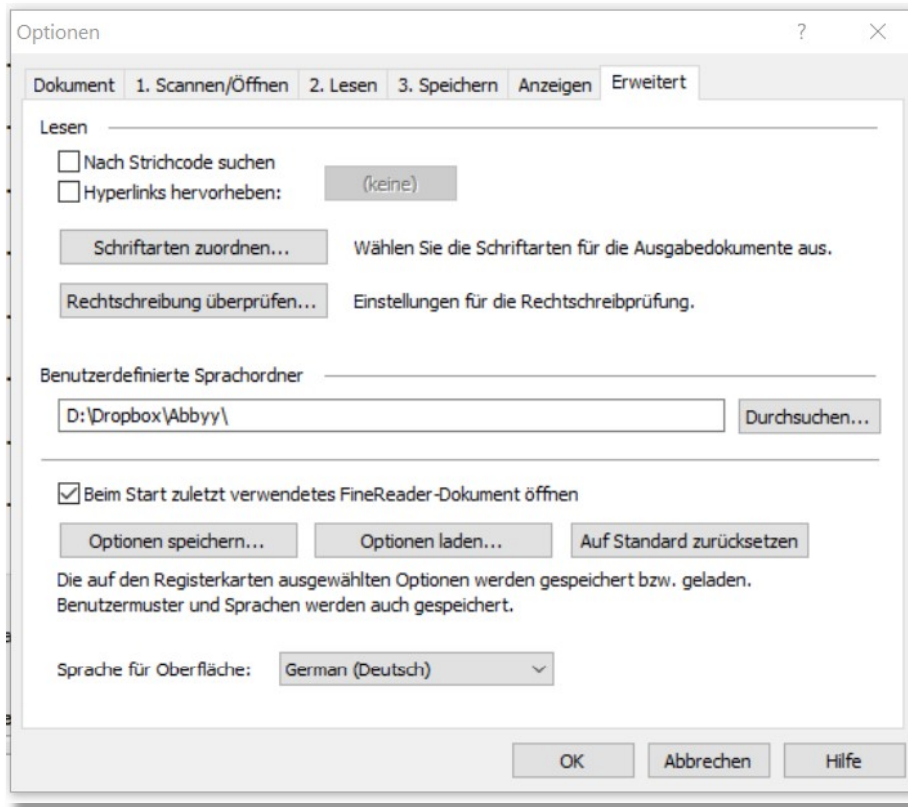


Abb. 23: FineReader-Option Erweitert

Hier gibt es Einstellungen für die Rechtschreibprüfung und für Speicherorte. Im **Benutzerdefinierte Sprachordner** habe ich das Wörterbuch abgelegt, das ich von dort exportieren kann. Abbyy legt dort seine eigenen Dateien ab wie z.B. vordefinierte Bereichsvorlagen, für die in Bearbeitung befindlichen Seiten und für die Optionen, die in diesem Kapitel besprochen werden.

5.1.7 Zwischenspeichern der OCR-Ergebnisse

Im Reiter **Erweitert** gibt es das Kästchen **Beim Start zuletzt geöffnetes FineReader Dokument öffnen**. Ein Häkchen bewirkt, dass nach Unterbrechung der Arbeit beim nächsten Start des FineReader die in Arbeit befindlichen Seiten wieder geladen werden. Man kann also das Bearbeiten der Seiten jederzeit unterbrechen und beim nächsten Start diese weiter bearbeiten.

5.2 Benutzerwörterbuch von FineReader

5.2.1 Benutzerwörterbuch als wichtiges Hilfsmittel zur Datenprüfung

FineReader kennzeichnet Wörter, die nicht bekannt sind, mit einer roten Unterstreichung. Dafür gibt es eine Reihe von verschiedenen Sprachen mit integrierten Wörterbüchern. Da es sich bei den Kirchenmatrikeln fast ausschließlich um Eigennamen und alten Berufsbezeichnungen handelt, ist die Verwendung des internen Wörterbuches mit den Flexionen wie Mehrzahl, Fall, Zeit etc. sinnlos bzw. sogar hinderlich. Sinnvoll ist nur ein eigenes Benutzerwörterbuch, in dem die als korrekt eingestuften Namen gespeichert werden.

Die rote Unterstreichung ist aber auch eine gute Möglichkeit, um durch das OCR-Programm Wörter bewusst hervorzuheben. Ich habe deshalb Wörter wie ledig, am, zu, auf, und nicht in das Wörterbuch aufgenommen. Ihre rote Unterstreichung macht mich darauf aufmerksam, dass z.B. nach dem Wort ledig immer ein Beistrich zur Worttrennung gesetzt sein muss oder dass bei Ortsangaben wie auf der Edt oder zu Tra-berg etc. diese unnötigen Zusätze zu löschen sind. Mit und werden oft 2 Berufe verbunden, die ebenfalls durch einen Beistrich zu trennen sind.

5.2.2 Benutzerwörterbuch erstellen

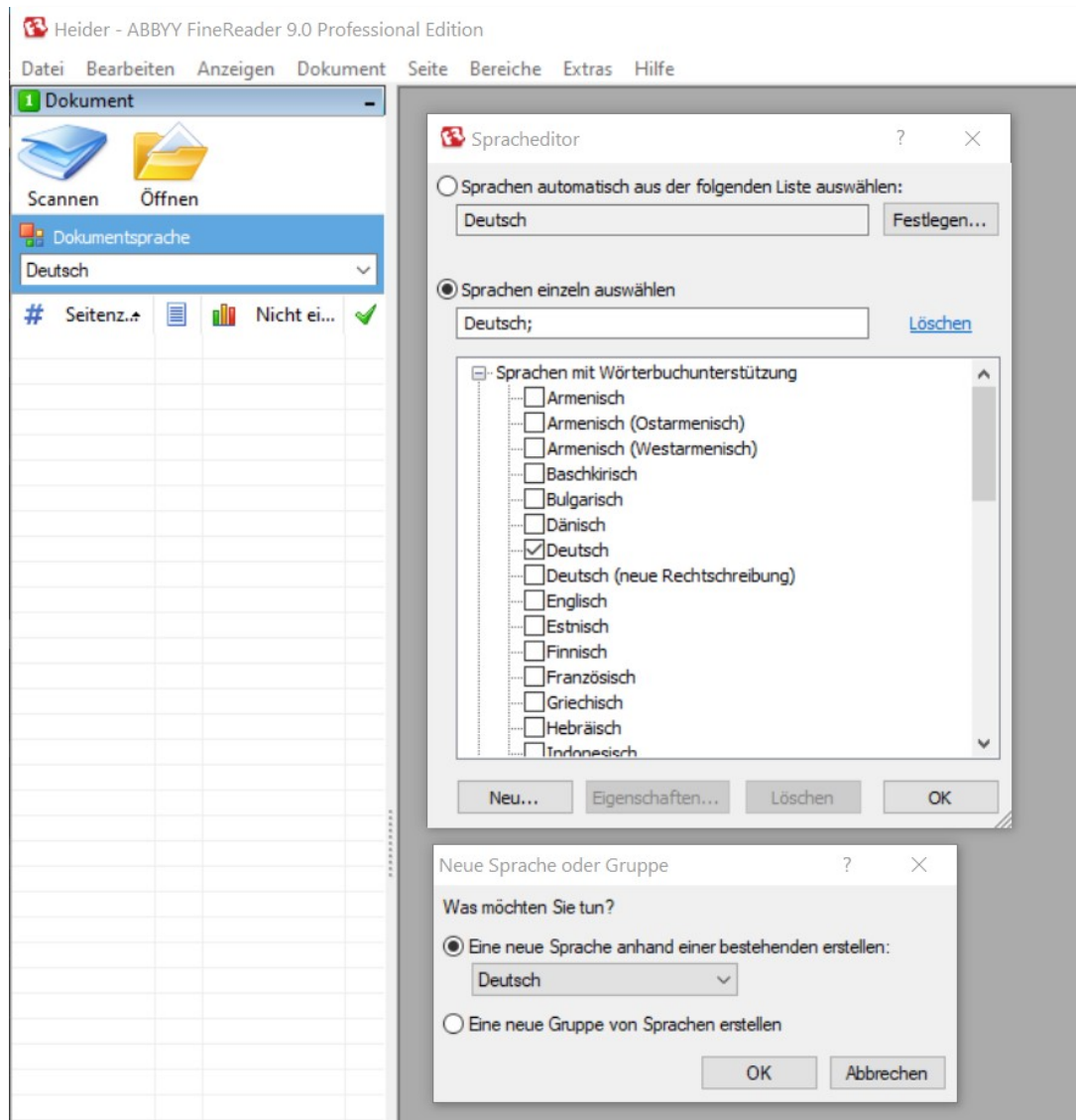


Abb. 24: FineReader-Option Spracheditor

Mit dem Einrichten eines eigenen Wörterbuches hatte ich immer wieder Schwierigkeiten und ich hoffe, dass ich nachfolgend den Ablauf richtig beschreibe (ich habe sehr lange herumexperimentiert, bis ich einen Weg gefunden habe).

Im Dokumentfenster von ABBYY FineReader ist am Anfang als **Dokumentsprache Deutsch** zu sehen. Mit **Extras/Spracheditor...** kann man eine oder mehrere andere Sprachen auswählen, oder mit **Neu...** neue Sprachen anlegen. Im nächsten Fenster ist **Eine neue Sprache anhand einer bestehenden erstellen:** auszuwählen. Ich habe Deutsch gewählt. Damit wird aber nur die alphabetische Sortierfolge festgelegt.

Im nächsten Fenster **Spracheigenschaften** gibt man unter **Sprachenname** einen geeigneten Namen an. Ich habe **Kirchenbuch** gewählt und als Wörterbuch **Benutzerwörterbuch**. Damit hat man ein neues, leeres Wörterbuch ohne Integration erstellt, das man jetzt im Dokumentfenster von ABBYY FineReader unter **Dokumentsprache** auswählen kann. Mit **Extras/Wörterbuch anzeigen...** sollte man das eben erstellte Benutzerwörterbuch auswählen und sicherstellen, dass das Kästchen für **Microsoft Word-Wörterbuch verwenden** **abgewählt** ist.

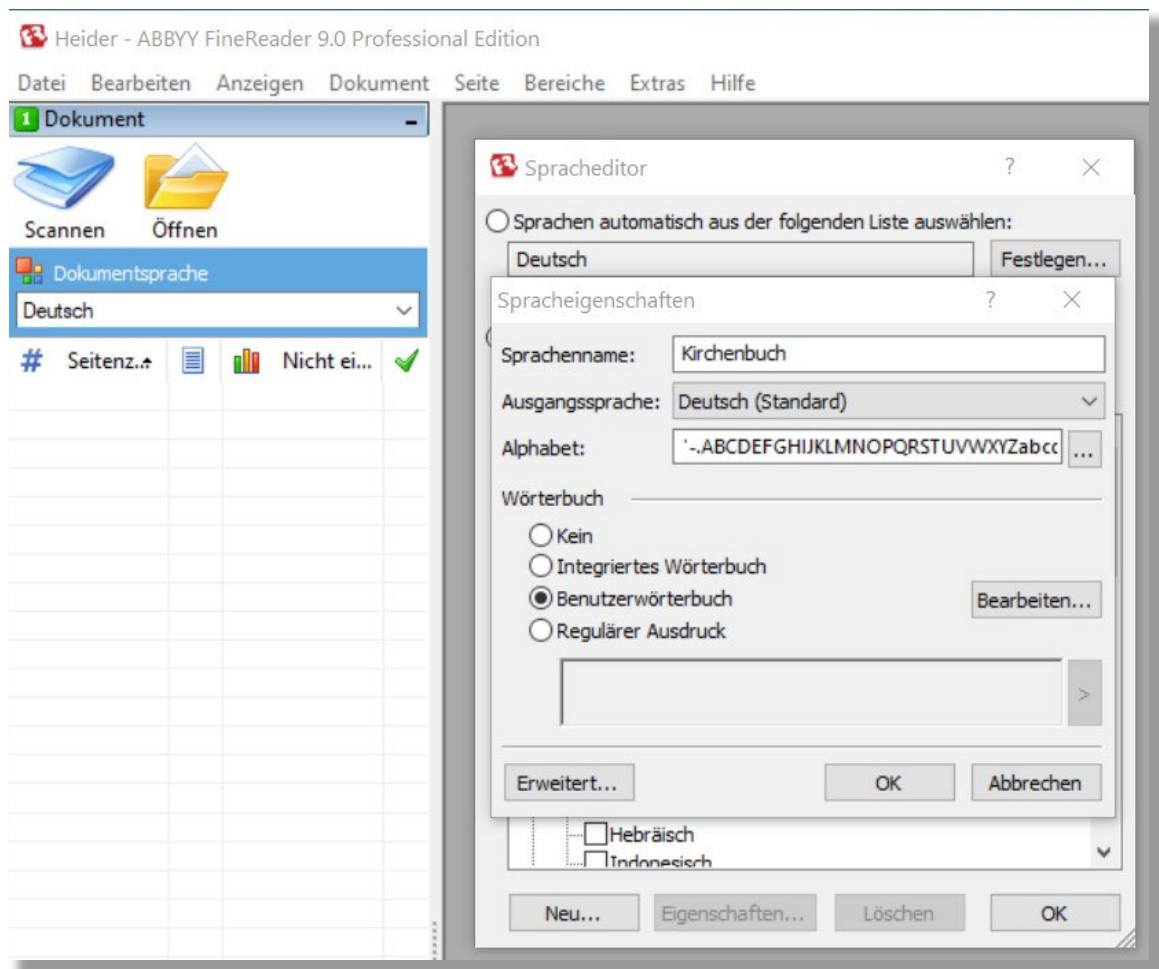


Abb. 25: FineReader: Neues Benutzewörterbuch anlegen

5.2.3 Neue Wörter im Benutzerwörterbuch speichern

Beim Auftreten neuer Begriffe ist zu überlegen, ob sie in das Benutzerwörterbuch übernommen werden sollen. Bei Standardnamen und -berufen ist das eindeutig. So wird man z.B. sehr bald die wichtigsten Vornamen und Berufe erfasst haben. Bei Familien- und Ortsnamen gibt es aber oft sehr viele Varianten und es ist manchmal nicht leicht zu entscheiden, ob eine Variante nur ein Schreibfehler in der Matrikel oder bei Heider ist. Auch sollte man prüfen, ob eine Variante vielleicht nur ein einmaliger Ausreißer ist oder doch öfter vorkommt. Je mehr Varianten im Wörterbuch gespeichert sind, desto größer die Gefahr, dass fehlerhaft geschriebene Namen nicht als Fehler angezeigt werden. Werden umgekehrt durch die OCR-Erkennung zu viele Begriffe durch die rote Unterstreichung markiert, besteht wieder die Gefahr, dass die Markierungen durch ihre Vielzahl nicht mehr ins Auge springen und übersehen werden. Außerdem hat das Wörterbuch Einfluss auf die OCR-Erkennung. Manchmal übersetzt der FineReader ein Wort in eines, das er im Wörterbuch findet, obwohl in der Registerseite etwas Anderes steht.

Ein Mangel des Benutzerwörterbuches ist die Speicherung der Begriffe ohne Großschreibung. Wird z.B. der Name Georg ins Wörterbuch aufgenommen, steht dort nur georg. Dadurch sind für den FineReader georg und Georg gültige Schreibweisen. Meine Macros prüfen daher zumindest den ersten Buchstaben des Namens und in der Namens-Spalte und erkennen diesen Fehler.

5.2.4 Benutzerwörterbuch exportieren/importieren

Namen sind oft sehr ortsspezifisch und kommen dann nur in bestimmten Pfarren vor. Es ist daher sinnvoll, die gespeicherten Begriffe von Zeit zu Zeit zu überprüfen und ausgefallene Begriffe und Schreibweisen wieder aus dem Wörterbuch zu löschen und damit dessen Umfang zu reduzieren. Leider werden die Begriffe im Fenster des Wörterbuches mit einem relativ kleinen Font angezeigt, was mir das Bearbeiten des Wörterbuches

ches schwermacht. Außerdem ist das Aufsuchen eines Begriffes nur durch Verschieben des seitlichen Schiebers möglich und nicht z.B. durch Eingabe der ersten Buchstaben. Man kann sich daher nur sehr mühsam an den gesuchten Begriff herantasten. Wenn man umfangreichere Änderungen am Wörterbuch vornehmen möchte, ist es daher einfacher, dieses zu exportieren und extern zu bearbeiten. Beim Exportieren wird eine Datei mit der dreistelligen Erweiterung **txt** gespeichert, die man mit KEDIT bearbeiten kann. Das veränderte Wörterbuch kann man mit der Importfunktion wieder zurückspeichern. Zuvor muss man aber das Benutzerwörterbuch im FineReader löschen, sonst tauchen die extern gelöschten Begriffe wieder auf.

In der Abbildung sind die 3 Schritte mit den zugehörigen Fenstern nebeneinander in einem einzigen Bild dargestellt.

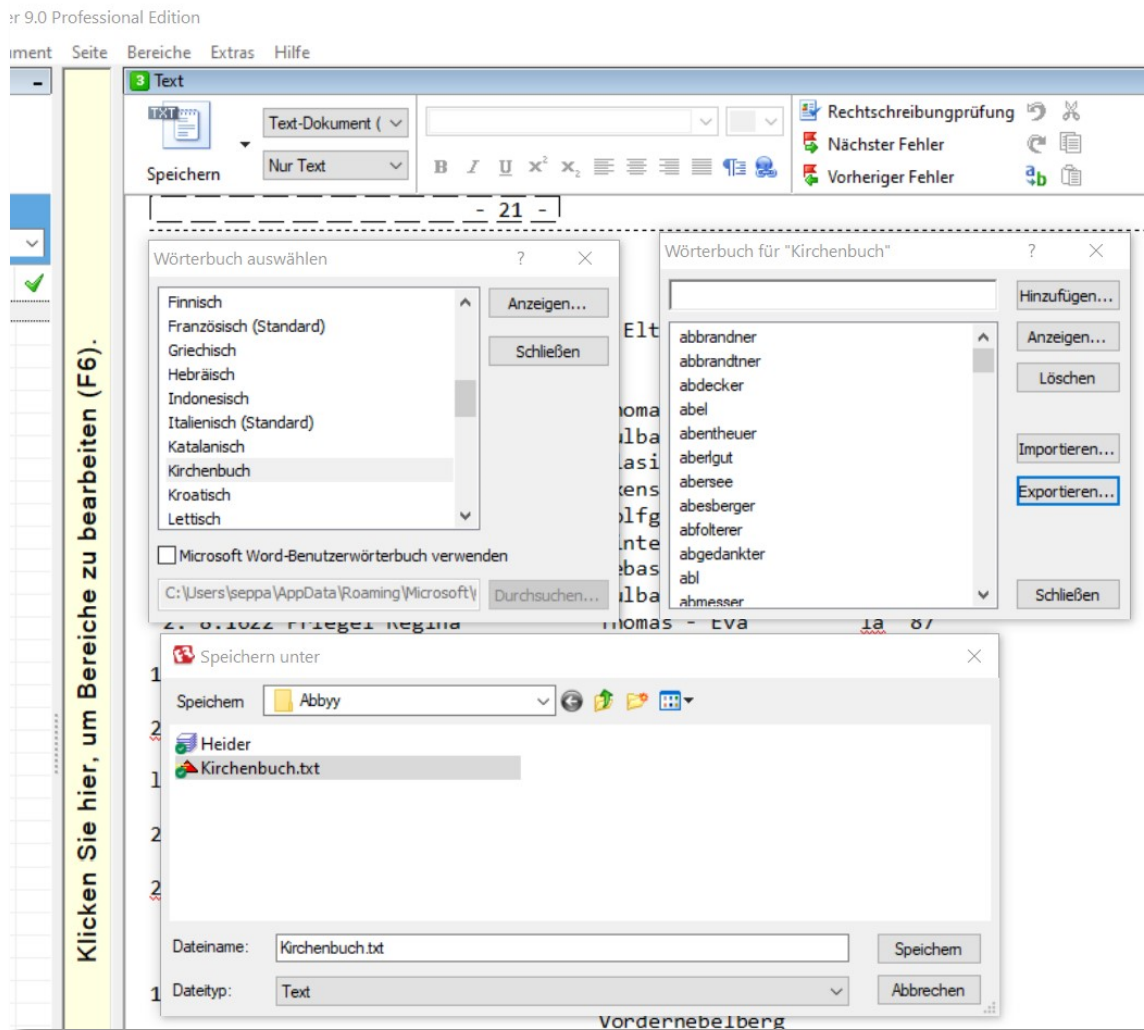


Abb. 26: Exportieren und importieren des Benutzerwörterbuches

6 KEDIT Texteditor: Installation und Editierfunktionen

6.1 Installation der KEDIT-Software und der Verarbeitungs-Programme (Macros)

KEDIT ist eine Nachbildung des IBM-XEDIT-Editors im VM-Betriebssystem von IBM Großrechnern. Es ist ein kostenpflichtiges Produkt von der Firma Mansfield Software Group, Inc. Auf der Homepage des Herstellers gibt es umfassende Hinweise zum Produkt: <http://www.kedit.com>. Von dieser Seite kann das Herunterladen von KEDIT gestartet werden. Auch die Dokumentationen (*User's Guide und Reference Manual*) können von dort heruntergeladen werden. Ich habe von der Support Group oft Unterstützung bekommen und kann KEDIT offiziell kostenlos für das Projekt verwenden.

KEDIT sucht nach den von Benutzern entwickelten Macros in verschiedenen Verzeichnissen. Das Standard-Verzeichnis dafür ist C:\Benutzer*user-name*\Dokumente. Zur Weitergabe der Macros habe ich diese in meiner Dropbox gespeichert und sie können von dort mit dem Link https://www.dropbox.com/scl/fo/w6ipxtwign4b9lj8sr819/AMESTIYFAM_Zhnr9TbY6mr8?rlkey=uxp6f3xm95ncwhy1o7d8l3yx8&dl=0 Heruntergeladen und in das oben angegebene Verzeichnis gespeichert werden (bitte den Link in den Browser kopieren).

6.2 Die wichtigsten Editor Zeilenbefehle

Das Formatieren und Prüfen einer Registerseite mit den von mir entwickelten KEDIT-Macros wurde bereits beschrieben und damit kommt man in der Regel auch aus. Es gibt aber noch eine Reihe von Editorbefehlen zum Löschen, Verschieben, Kopieren von Texten und ganzen Zeilen und Textblöcken, auf die ich hier für den Bedarfsfall eingehe.

Durch die Einstellungen im Profile für die Arbeit an unserem Projekt wird das KEDIT-Fensters mit der sog. prefix area dargestellt (Inhalte und Aussehen des Edit-Fensters lassen sich konfigurieren). Die prefix area ist die 5-stellige Nummer am Beginn der Editor-Zeilen, wie sie in den verschiedenen KEDIT-Beispielen dieser Doku zu sehen ist. In einem oder mehreren dieser prefix area kann man Befehle eingeben, die nach Drücken der **PF12-Taste** ausgeführt werden.

6.2.1 Einfache Befehle

- **a** (add) oder **i** (insert): eine Leerzeile wird eingefügt
- **d** (delete): eine Zeile wird gelöscht
- **c** (copy) Zeile wird vor die Zeile mit einem **p** (previous) oder nach der Zeile mit **f** (following) kopiert
- **f** (following) kopierte oder verschobene Zeile wird danach eingefügt
- **m** (move) Zeile wird vor Zeile mit **p** oder nach der Zeile mit **f** verschoben
- **p** (previous) kopierte oder verschobene Zeile wird davor eingefügt
- **l** (lowercase) alle Buchstaben werden auf Kleinbuchstaben gesetzt
- **u** (uppercase) alle Buchstaben werden auf Großbuchstaben gesetzt
- **„** Zeile wird dupliziert
- **<** Zeile wird um eine Stelle nach links verschoben
- **>** Zeile wird um eine Stelle nach rechts verschoben

Das Verschieben mit den **< >** Befehlen betrifft immer die ganze Zeile. Um einen Teil des Textes in einer oder in mehreren Zeilen zu verschieben verwendet man die Block-Funktion. Siehe *Arbeiten mit Textblöcken* auf Seite 42.

6.2.2 Befehle mit Zahlenangaben

Vor oder nach den Befehl **a**, **i**, **d**, **c**, **m**, **l**, **u**, **„**, kann eine Zahl angegeben werden. Damit wird die Anzahl der Zeilen angegeben, für die der Befehl wirksam sein soll. bzw. um wieviel Stellen ein Text verschoben werden

- **a3** füge 3 Zeilen ein
- **d4** lösche die nächsten 4 Zeilen

Beim Verschieben Textdefiniert die Zahl die Anzahl zu verschiebenden Stellen:

- **>5** verschiebe den Text der Zeile um 5 Stellen nach rechts

6.2.3 Verdoppelung der Befehle

Als Alternative zu den Zahlenangaben kann man auch Anfang und Ende der zu bearbeitenden Zeilen angeben. Dazu gibt man in der prefix area der ersten und der letzten Zeile den verdoppelten Befehles ein:

cc, dd, ll, mm, uu, <<, >>

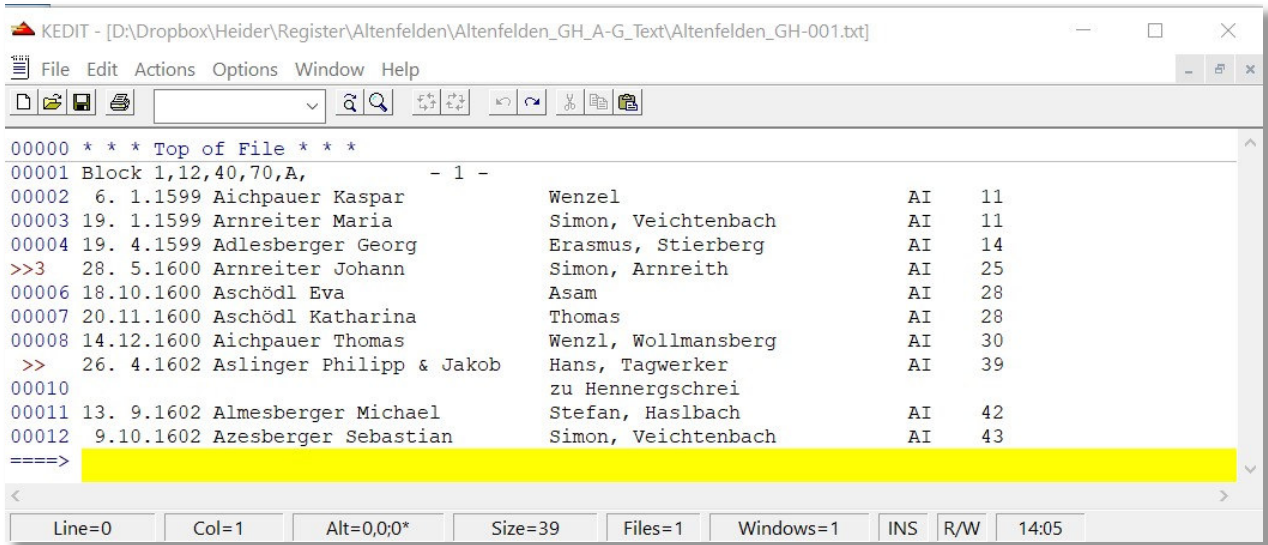


Abb. 27: Auswahl von mehreren Zeilen

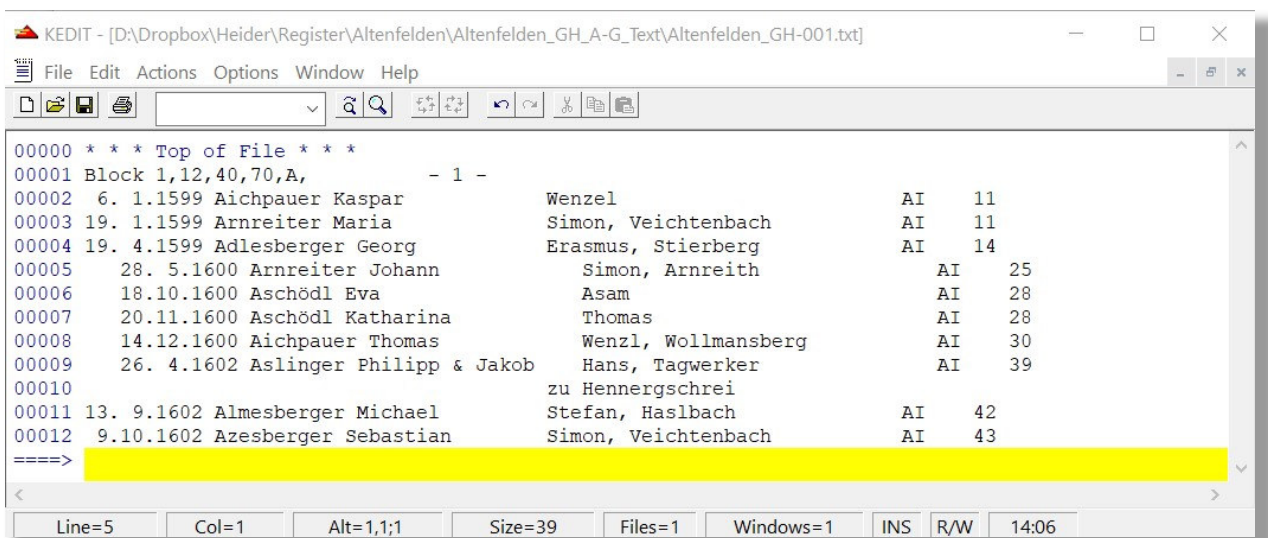


Abb. 28: verschiebe Zeilen 5 bis 9 um 3 Stellen nach rechts

6.3 Arbeiten mit Textblöcken

KEDIT kann auch mit Textböcken arbeiten. Einen Block wählt man aus, indem man den Cursor auf den ersten Buchstaben des Blockes stellt und dann die Maus mit gedrückter Alt-Taste nach rechts über den auszuwählenden Bereich zieht. Nun gibt es eine Reihe von Manipulationen, die sich auf einen ausgewählten Textblock beziehen. Für sie gibt es kleine Icons in der bottom toolbar. Die bottom toolbar wird bei Aufruf von

Sterbematriken automatisch aktiviert, kann aber auch durch Drücken der **F2**-Taste (toggle bottom toolbar) am unteren Rand des Fensters eingeblendet werden. Sobald ein Textblock ausgewählt wird, werden die Icons aktiviert. Die folgenden zwei Beispiele sollen das Arbeiten mit Textblöcken veranschaulichen.

6.3.1 Textblock in den Kommentar verschieben

In der Registerseite 20 des Taufbuches von St. Johann am Wimberg hat J. Heider einen Originalltext aus der Taufmatrikel übernommen. Durch Auswahl des Textblockes und Ziehen desselben mit der Maus lässt sich der Text als Ganzes an die Position des Kommentares verschieben. Nach Drücken der Eingabetaste werden dann auch wie bei allen anderen Texten mehrfache Leerzeichen entfernt.

00041	26. 5.1730 Plank Johann	Johann, Glasmacher	III 151
00042			
00043			
00044			
00045			
00046			

aus Frankfurth am Main gebürtig, laut seiner Matrimonial-Attestation in rechtmässiger Ehe, Vagus. Mater Anna Maria von Stift Fulda im frankenland gebürtig.

Abb. 29: Textblock von Namens- und EBO-Spalte

26. 5.1730 Plank Johann	Johann, Glasmacher	III 151	aus Frankfurth am Main gebürtig, laut seiner Matrimonial-Attestation in rechtmässiger Ehe, Vagus. Mater Anna Maria von Stift Fulda im frankenland gebürtig.
-------------------------	--------------------	---------	---

Abb. 30: in den Kommentarbereich verschieben

6.3.2 Verschieben der Altersangaben in den Sterberegistern

Das Verschieben von Textblöcken ist vor allem bei den Sterberegistern eine sehr nützliche Funktion. Das in der folgenden Abbildung gezeigte Beispiel ist die 2. Seite des Sterberegisters von Aigen im Mühlkreis. Unter *tomus-Block und Altersangabe* beim Sterberegister auf Seite 25 wird erklärt, warum dieser Textblock verschoben wird. In diesem Beispiel wurde der Cursor auf die Position 54 gesetzt, der Textblock mit der Maus ausgewählt und mit der Funktion **Shift Block Right** um 3 Stellen nach rechts verschoben.

```

00000 * * * Top of File * * *
00001   Tag      Name                Eltern, Beruf, Ort      Tom.
00002
00003
00004 18.10.1674 Aichinger Stefan      Aigen                   II 247
00005
00006 20. 8.1674 Autengruber Michael  Rudolfing              II 247
00007
00008 13. 5.1675 Autengruber Anna      11 248
00009
00010 11. 9.1675 Autengruber Martha    Totengräber, Wächte   r II 248
00011
00012 4. 6.1678 Amprosin Maria        Praitenstein           II 254
00013
00014 6. 6.1682 Almesberger Kind          Florian, Schindlau     II 261
00015
00016 23. 3.1684 Autengruber Andreas    Rudolfing              11 265
00017
00018 2.12.1688 Almesberger Adam        Aigen                   II 274
00019
00020 14. 9.1715 Almesberger Kaspar    Schindlau              6o III 496
00021
00022 22. 6.1716 Almesberger M. Magdalena Schindlau             5 III 498
00023
00024 9. 6.1718 Almesberger Maria      Wurmbrand              6 III 5o5
00025
00026 6.12.1719 Asanger Johann            Knecht zu Aigen       3o III 5o9
00027
00028 5. 5.1723 Aigner Lukas                Landgerichtsdienere   6o IV 589
00029                zu Aigen
00030 24. 1.1725 Almesmüller Philipp    wurmbrand              7o IV 595
00031
00032 26. 8.1726 Almesberger Magdalena Hans Georg, Schindl  au IV 6o3
00033                15M
00034 9. 7.1787 Amgner Hans Beong      ledig, Aigen           28 IV 6o9
00035
00036 6. 2.1729 Asanger Maria        Michael, ledig        11W IV 613
00037                zu Unterneudorf
00038 7. 4.1729 Almesberger Gertrude    Hans Georg             7 IV 615
00039
00040 22.12.1729 Asanger Margarete          Holzhäuseln           6o IV 619
00041
00042 1o. 3.173o Aigner Eva Maria        Landgerichtsdienere   71 IV 621
00043                zu Aigen
00044 17. 9.173o Aigner M. Katharina    Kaspar, Langerichts  -7T IV 624
00045                diener zu Aigen
00046 9. 7.1731 Andexlinger Maria      Rudolfing             51 IV 627
00047
00048 29. 9.1731 Aigner M. Magdalena          Kaspar, Landgericht  s- IV 628
00049                diener zu Aigen    low
00050 27.12.1731 Aigner Anna Maria          Landgerichtsdienere   24 IV 629
00051                zu Aigen
00052 7. 2.1732 Asanger Thomas        Johann, Zaglau        7W IV 63o
00053
00054 28, 3.1732 Almesberger Rosina    Hans Georg, zu       4 IV 63o
00055                Schindlau
00056 * * * End of File * * *

```

Abb. 31: Textblock
verschieben

7 Abbildungsverzeichnis

Abb. 1: Seite eines Taufregisters	7
Abb. 2: Seite eines Sterberegisters	8
Abb. 3: Öffnen von 10 Registerseiten.....	11
Abb. 4: Auswahlrahmen erstellen	13
Abb. 5: Tag außerhalb des Erkennungsrahmens.....	14
Abb. 6: lesen und bearbeiten der Registerseiten.....	15
Abb. 7: OCR Textänderungen rückgängig machen	17
Abb. 8: Ausschnitt aus dem KEDIT-Profil	19
Abb. 9: Die Seite 171B wurde eingeschoben, weil Heider die Nr. 171 zweimal vergeben hat	22
Abb. 10: alternative Schreibweise des Familiennamens.....	23
Abb. 11: Textblock mit Alter und tom-Block nach rechts verschoben	26
Abb. 12: EBO-Information und Altersangabe entflechtet	27
Abb. 13: formatierte Seite des Sterberegisters mit Altersangaben.....	28
Abb. 14: formatierte Registerseite mit eingefügter Blockzeile	30
Abb. 15: Vornamen in die Folgezeile verschoben	31
Abb. 16: Meldung im Kommentarfeld	32
Abb. 17: Rückgängig machen von Änderungen in KEDIT	33
Abb. 18: FineReader-Option Dokument.....	34
Abb. 19: FineReader-Option Scannen/Öffnen	35
Abb. 20: FineReader-Option Lesen	35
Abb. 21: FineReader-Option Speichern	36
Abb. 22: FineReader-Option Anzeigen	36
Abb. 23: FineReader-Option Erweitert	37
Abb. 24: FineReader-Option Spracheditor	38
Abb. 25: FineReader: Neues Benutzewörterbuch anlegen.....	39
Abb. 26: Exportieren und importieren des Benutzewörterbuches.....	40
Abb. 27: Auswahl von mehreren Zeilen	42
Abb. 28: verschiebe Zeilen 5 bis 9 um 3 Stellen nach rechts.....	42
Abb. 29: Textblock von Namens- und EBO-Spalte	43
Abb. 30: in den Kommentarbereich verschieben.....	43
Abb. 31: Textblock verschieben	44