

Digitale Prüfung

Thema: Schwierige Dateien für die Datenanalyse- Probleme und Lösungsstrategien

In diesem Newsletter wenden wir uns ausschließlich dem operationalen Thema „Schwierige Dateien“ zu, mit welchem uns die Prüfungspraxis oft konfrontiert. Hierzu erreichen uns auch regelmäßig zahlreiche Fragen von Kolleginnen und Kollegen.

Prüferische Datenanalysen sind ein fester Bestandteil im Werkzeugkasten der prüferischen Urteilsbildung. Unabhängig von den hierzu eingesetzten Techniken (Prüfsoftware, SQL- und QBE-Datenbankabfragen, Berichtsgeneratoren, grafische Dashboards oder KI-Verfahren) müssen zunächst die *erforderlichen Daten bereitstehen*. Hierbei stellt der Datenaustausch zwischen Ausgangs- und Analysesystemen in der Regel kein Problem dar, wenn auf ein bekanntes Datenbankformat oder eine gängige Schnittstelle (z.B. Excel, Access, ODBC) zurückgegriffen werden kann. Oft jedoch müssen zahlreiche prüfende Stellen auf *Daten unterschiedlichster Vorverfahren* zugreifen, die keinen optimalen Datenaustausch ermöglichen. Vielfach ist die Nutzung hier bereitstehender Datenquellen (*Text-, Druck- PDF- oder diverse Excelformate*) mit zahlreichen Fiktionen verbunden, die zunächst erkannt und anschließend eingeordnet werden müssen, in Probleme, die mit vertretbarem Aufwand eine Lösung ermöglichen und solche, für welche sich die Anstrengungen einer Nachbearbeitung nicht lohnen. Hierzu finden sich im Folgenden einige Hinweise.

1 Geeignete Werkzeuge als solide Grundausstattung

Kaum jemand mag sich einen Heizungsmonteur vorstellen, der sich ohne geeignetes Werkzeug mit der beschädigten Gastherme beschäftigt. Es ist daher erstaunlich, mit welchem Enthusiasmus sich veritable Revisorinnen und Revisoren ausschließlich mit Windows-Bordmitteln einem Problembereich zuwenden, der (außer praktischer Erfahrung) professioneller Tools bedarf. Zwar können viele Unzulänglichkeiten in Ausgangsdateien mit überschaubarem Aufwand behoben werden, in schwierigeren Fällen erweist sich dieses Vorgehen jedoch nicht nur als aufwendig, sondern kann die Problemzonen sogar noch ausweiten. Dieses ist bedauerlich, denn gute und praxisbewährte Tools sind - im Verhältnis zu deren Nutzen - preiswert erhältlich. Wir stellen daher in einem ersten Schritt einige bewährte Werkzeuge (ohne Anspruch auf Vollständigkeit) vor:

- **Unverzichtbar: Ein geeigneter Texteditor**

Die Problemanalyse und -lösung in großen Ausgangsdateien erfordert Editoren, die mit Massendaten in unterschiedlichsten Formaten (Text, Unicode, HTML, XML, Binär etc.) umgehen können. Dieses gilt sowohl für bildliche, formatierte und hexadezimale Darstellungen wie für Änderungen an solchen Daten. Windows-Zubehör zählt eher nicht dazu! Internetrecherchen führen hingegen schnell zu geeigneten Lösungsalternativen, die von Freeware bis zu kostengünstigen Entwicklungen (z.B. UltraEdit) reichen. Ein gerne verwendeter Editor, welcher mit großen Dateien in unterschiedlichsten Formaten umgehen kann, ist Bestandteil des verbreiteten Dateimanagers „Total Commander“ (einmalig ca. 40 Euro).

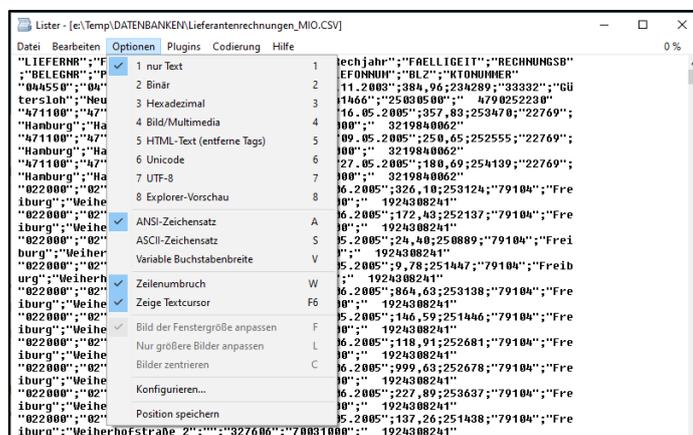


Abb. 1 Editor des Total Commanders mit unterschiedlichen Bearbeitungsformaten

Die empfehlenswerte Software verfügt über zahlreiche sinnvolle Funktionen und hat sich in praktischen Analyseprojekten vielfach bewährt. Zu weiteren Editoren (zum UltraEdit) erhält man im Internet umfassende Informationen.

- **Bei PDF-Formaten: Dateibetrachter mit Transformationsmöglichkeiten**

Ein Austausch von Informationen über das verbreitete PDF-Format (Portable Document Format) zum Zwecke der Datenanalyse ist regelmäßig mit zahlreichen *technischen* Problemen verbunden. Diese reichen von vektorbasierten PDF-Inhalten, verschlüsselter Speicherung bis gescannten PDF-Dokumenten mit schlechter Qualität. Die aufgeführten Problembereiche stehen neben *möglichen inhaltlichen Zumutungen* (z.B. druckaufbereitete Formatierungen) die sich in einem zweiten Schritt hinzugesellen können.

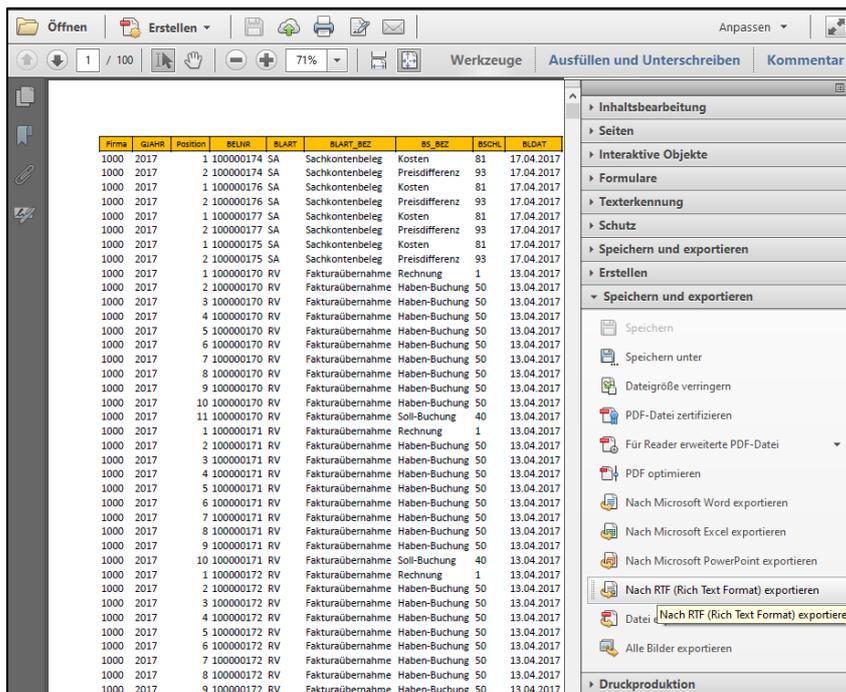


Abb. 2 Adobe Acrobat Pro mit unterschiedlichen Transferoptionen

Zur Lösung dieser technischen Probleme eignen sich spezialisierte PDF-Tools, die u.a. in der Lage sind, PDF-Dateien in andere Formate (RTF, Word, Excel) zu überführen. Die Alternativen reichen von „Adobe Acrobat Pro“ (nicht der frei erhältliche Acrobat Reader), welches für ca. 27 Euro monatlicher Lizenzgebühr angeboten wird, bis zu leistungsfähiger Freeware. Einen Überblick vermittelt u.a. die nachfolgende Internetsseite: „<https://www.top5pdftools.de>“.

- **Für die Aufbereitung: Spezielle Transformations- und Importprogramme**

Der Import „schwieriger“ Dateien ist eigentlich eine Domäne tradierter Prüfsoftware (ActiveData, ACL und IDEA). Deren Anwendung im Prüfungsalltag zeigt jedoch, dass die Fähigkeiten, mit wirklich schwierigen Importsituationen umzugehen, sehr unterschiedlich ausgeprägt sind.

- ACL-Prüfsoftware

ACL verfügt über sehr leistungsfähige Übernahmemöglichkeiten, erfordert jedoch gute Kenntnisse zur Handhabung hierfür bereitstehender Funktionen. Für kritische Druck- und PDF-Daten fehlt überdies eine Vorschaufunktion zur Beurteilung von Zwischenergebnissen.

- IDEA-Prüfsoftware

IDEA bedient sich bei der Übernahme kritischer Druck- und PDF-Daten mit dem angegliederten Report Reader eines extern integrierten Programms. Bei Standardproblemen offeriert dieser gute Lösungsoptionen. Für wirklich schwierige Import-Situation ist er jedoch weniger geeignet.

- ActiveData-Prüfsoftware

ActiveData stützt sich bei der Übernahme von Ausgangsdaten hauptsächlich auf zugehörige Excel-Funktionalität. Diese ermöglicht den Zugriff auf alle gängigen Datenbank-, Text- und Excel-dateien, nicht jedoch die einfache Behandlung von PDF- und Druckdateien.

Angesichts der überschaubaren Lösungskompetenz gängiger Prüfsoftware hat sich bei praktischen Anwendungen die spezialisierte Software „Import Wizard“ als *beeindruckend leistungsfähiges, universelles, leicht handhabbares und effizientes Werkzeug* erwiesen, welches in einer einfachen Form alle gängigen Datenformate behandelt und selbst bei kritischsten Ausgangsdateien noch zu guten, analysefähigen Ergebnissen führt.

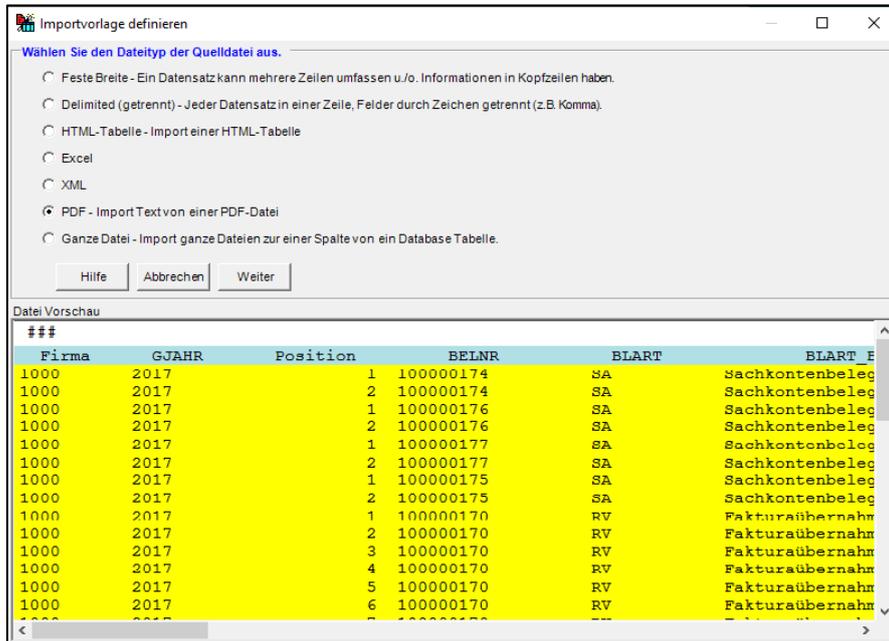


Abb. 3 ImportWizard als universeller Problemlöser für unterschiedlichste Dateiformate

Hierbei beschäftigt es sich nicht nur mit den Problemen von Druck- und PDF-Daten, sondern ebenso von Excel-, Text-, HTML- und XML-Formaten. Mit einem günstigen Preis-Leistungsverhältnis (einmalig 199 €) vereinfacht es selbst für die Anwender tradierter Prüfsoftware den Datenübernahmeprozess wesentlich. Als Praxisempfehlung beziehen wir unsere nachfolgenden Lösungshinweise insbesondere auf dieses Tool.

- **Für externe Abschlussprüfer: Der GoBD-Importer**

Entgegen weit verbreiteter Auffassung sind GoBD-Dateien bei Datenübernahmen durchaus nicht immer problemfrei. Da sie nicht aus „einem Guss“ bestehen, sondern aus einer Kombination von eigentlichen Datendateien (häufig ein Textformat mit variablen Feldlängen und Feldtrenner), zugehörigen Satzbettbeschreibungen im XML-Format (Index.XML) und einer Schema-Datei (Suffix „DTD“) mit allgemeinen Formatierungs- und Bindungsinformationen liegen mögliche Problemfelder auf unterschiedlichen Ebenen.

Unzutreffende Adressangaben zu den eigentlichen Daten beeinträchtigen den GoBD-Import ebenso wie unzutreffende Satzbettangaben, falsche Formatierungsinformationen, die Speicherung in einer unrichtigen Kodierung oder die fehlerhafte Berücksichtigung von Feldtrennern in Datendateien.

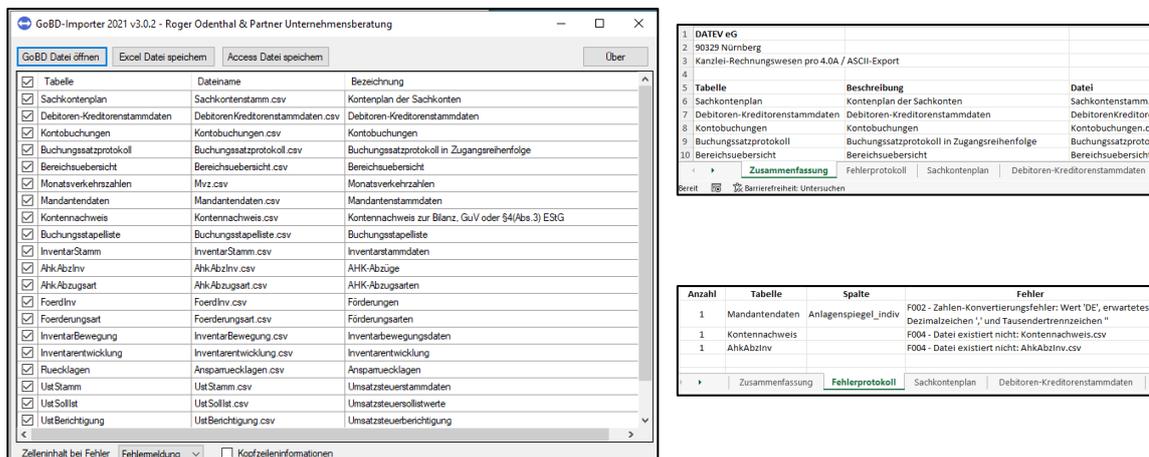


Abb. 4 GoBD-Importer mit Datei- und Ergebnisübersicht sowie Fehlerprotokoll

Entsprechende Fehler entdeckt der GoBD-Importer (einmalig 149 Euro) weitgehend automatisch, bevor er sämtliche GoBD-Daten für die weitere Analyse auf Knopfdruck in eine Excel- oder Accessdatei übernimmt. Für professionelle Datenanalysten insbesondere aus dem Umfeld der externen Jahresabschlussprüfung kann dieses Werkzeug ebenfalls empfohlen werden.

Unsere unter Effizienz- und Kostengesichtspunkten zusammengestellte Toolbox für schwierige Dateien beinhaltet hiernach den Total Commander (mit integriertem Editor), einen PDF-Editor (hier Acrobat Pro), den ImportWizard als Passepartout für die einfache und automatisierte Transformation schwieriger Dateien sowie den GoBD-Importer für das Handling von Buchhaltungsdaten. Die hier aufgeführte Profiausrüstung ist für unter 500 € (einmalige Gebühr) erhältlich und rechnet sich nach praktischer Erfahrung für jeden Anwender (unabhängig von der verwendeten Prüfsoftware) bereits im Zusammenhang mit dem erstmaligen Einsatz.

2 Ausgewählte Einzelprobleme und Lösungsmöglichkeiten

Nachfolgend geschilderte Einzelprobleme und deren Behandlungsoptionen sind jeweils von den einzelnen überlassenen Datenformaten abhängig:

- **Textdateien mit festen Feld- und Satzlängen**

Es handelt sich um eine eher seltene Form der Datenüberlassung, da leere Inhalte als Daten transportiert werden, was regelmäßig für ein umfangreiches Datenvolumen sorgt. Der Transfer von SAP-Tabellen in unkonvertiertem Textformat ist ein Beispiel. Probleme können auftauchen, wenn einzelne Felder oder Datensätze (sofern sie ohne Satzbegrenzer „CR / LF“ angeliefert wurden) fehlende Stellen aufweisen. Hieraus folgen bei allen nachfolgenden Datensätzen sinnentstellenden Verschiebungen zu Feldinhalten. Bei Dateien mit Satzbegrenzer treten diese fehlerhaften Inhalte lediglich innerhalb einzelner betroffener Datensätzen auf.

Lieferantennu	Rechnungsb	Rechnungsdatum	Faelligkeit	Belegnummer	Name	Ort	PLZ	Strasse	Vorwahl	Telefon	Konto	BLZ
1	LIEFERNR	0,00			LIEFNAME	ORT	PLZ	STRASSE	VORWAHL	TELEFONNUMKTO	NUMMER	BLZ
2	044550	384,96 31.10.2003	30.11.2003	234289	A.B.C.	Gütersloh	33332	Neuenkirchener Straße 55	05241	841466	4790252230	25030500
3	471100	357,83 16.04.2005	16.05.2005	253470	AKH FERRO	Hamburg	22769	Hafervweg 11	040	1391735	3219840062	50080000
4	4711	250,65		252555AK	H FERRO	mburg	22 769Ha	ferweg 11	04 0	1391735 32	19840062 50	080000
5	471100	180,69 27.04.2005	27.05.2005	254139	AKH FERRO	Hamburg	22769	Hafervweg 11	040	1391735	3219840062	50080000
6	022000	326,10 16.04.2005	29.06.2005	253124	ALBAN	Freiburg	79104	Weierhofstraße 2		327606	1924308241	70031000
7	022000	172,43 09.04.2005	29.06.2005	252137	ALBAN	Freiburg	79104	Weierhofstraße 2		327606	1924308241	70031000
8	022000	24,40 30.03.2005	30.05.2005	250889	ALBAN	Freiburg	79104	Weierhofstraße 2		327606	1924308241	70031000
9	022000	9,78 31.03.2005	30.05.2005	251447	ALBAN	Freiburg	79104	Weierhofstraße 2		327606	1924308241	70031000
10	022000	864,63 16.04.2005	29.06.2005	253138	ALBAN	Freiburg	79104	Weierhofstraße 2		327606	1924308241	70031000
11	022000	146,59 31.03.2005	30.05.2005	251446	ALBAN	Freiburg	79104	Weierhofstraße 2		327606	1924308241	70031000

Abb. 5 Datensatz mit fehlerhaften Inhalten (auch Beträge und Konten) durch fehlende Zeichen

Ein gängiger Grund (neben anderen) für dieses Problem ist der Wechsel von Zeichentabellen (Codepage) mit unterschiedlichem Ausgangs- (z.B. Unicode) und Zielformat (z.B. ANSI) ohne begleitende Transformation. Wird lediglich in das neue Format gespeichert und nicht während der Datenerzeugung bereits transformiert, so werden zumindest Umlaute unrichtig und mit mehreren (statt einem) Zeichen dargestellt, was die entsprechende Verschiebung zur Folge hat.

Lieferan	Rechnungsb	Rechdat	Faelligkeit	BelegNr	Name	Ort	PLZ	Strasse	Vorwahl	Telefon	Konto	BLZ
2	044550	384,96 31.10.2003	30.11.2003	234289	A.B.C.	GA'tersl oh	3333	2Neuenkirchener Straße 55	0524	841466	4790252230	25030500
3	471100	357,83 16.04.2005	16.05.2005	253470	AKH FERRO	Hamburg	22769	Hafervweg 11	040	1391735	3219840062	50080000
4	4711	250,65		2555AK	H FERRO	Ha mburg	22769Ha	ferweg 11	40	1391735 3	219840062	0080000
5	471100	180,69 27.04.2005	27.05.2005	254139	AKH FERRO	Hamburg	22769	Hafervweg 11	040	1391735	3219840062	50080000
6	022000	326,10 16.04.2005	29.06.2005	253124	ALBAN	Freiburg	79104	Weierhofstraße 2		327606	1924308241	70031000
7	022000	172,43 09.04.2005	29.06.2005	252137	ALBAN	Freiburg	79104	Weierhofstraße 2		327606	1924308241	70031000
8	022000	24,40 30.03.2005	30.05.2005	250889	ALBAN	Freiburg	79104	Weierhofstraße 2		327606	1924308241	70031000
9	022000	9,78 31.03.2005	30.05.2005	251447	ALBAN	Freiburg	79104	Weierhofstraße 2		327606	1924308241	70031000
10	022000	864,63 16.04.2005	29.06.2005	253138	ALBAN	Freiburg	79104	Weierhofstraße 2		327606	1924308241	70031000

Abb. 6 Datensatz mit fehlerhaften Inhalten durch unrichtige Zielformate (Zeichentabelle) der Ausgangsdatei

Der gängige Lösungsansatz bei Einzelfehlern besteht in der Verwendung eines Texteditors und der automatisierten Umwandlung (Suchen und Ersetzen) fehlerhafter Zeichen. Daneben kann eine Durchsicht in der Datei fehlerhafte Datensätze (zum Beispiel unrichtige Datensatzlänge) identifizieren und nachfolgende manuelle Korrekturen ermöglichen. Beides sollte vor dem Import durch Prüfsoftware erfolgen. Nach der Übernahme in Prüfsoftware ist eine zusätzliche Verprobung zu Abstimmsummen sowie zur Anzahl übernommener Positionen zu empfehlen, da in großen Dateien solche Fehler oft unbemerkt bleiben.

LIEFERNR	RECHNUNGSRECHNUNGSD	FAELLIGEIT	BELEGNRLIEFNAME	ORT	PLZ	STRASSE	VORWAHL	TELEFONNUMKTONUMMER	BLZ
044550	384,9631.10.2003	30.11.2003	234289A.B.C.	Gütersloh	33332	Neuenkirchener Straße 55	05241	841466 4790252230	25030500
471100	357,8316.04.2005	16.05.2005	253470AKH FERRO	Hamburg	22769	Hafervweg 11	040	1391735 3219840062	50080000
4711	250,6509.04.2005	09.05.2005	252555AKH FERRO	Hamburg	22769	Hafervweg 11	040	1391735 3219840062	50080000
471100	180,6927.04.2005	27.05.2005	254139AKH FERRO	Hamburg	22769	Hafervweg 11	040	1391735 3219840062	50080000
022000	326,1016.04.2005	29.06.2005	253124ALBAN	Freiburg	79104	Weierhofstraße 2		327606 1924308241	70031000
022000	172,4309.04.2005	29.06.2005	252137ALBAN	Freiburg	79104	Weierhofstraße 2		327606 1924308241	70031000
022000	24,4030.03.2005	30.05.2005	250889ALBAN	Freiburg	79104	Weierhofstraße 2		327606 1924308241	70031000
022000	9,7831.03.2005	30.05.2005	251447ALBAN	Freiburg	79104	Weierhofstraße 2		327606 1924308241	70031000
022000	864,6316.04.2005	29.06.2005	253138ALBAN	Freiburg	79104	Weierhofstraße 2		327606 1924308241	70031000

Abb. 7 Kontrolle und Korrekturmöglichkeit zu fehlerhaft übertragenen Ausgangsdaten mittels Editor

Achtung: Ein einfacher Wechsel der Zeichentabelle während des Imports (z.B. von Unicode nach ANSI) ist nicht möglich, wenn die Daten während des Downloads aus dem Ausgangssystem lediglich fehlerhaft gespeichert, aber nicht begleitend transformiert wurden. Falls lediglich Umlaute von der unzureichenden Handhabung betroffen sind, kann häufig eine spätere Korrektur (z.B. mittels neuer Rechenfelder und „Austauschfunktion“ in der Prüfsoftware) erfolgen.

- **Textdateien mit *variablen* Feld- und Satzlängen (Delimited-Format)**

In diesem Textformat Analysedateien häufig bereitgestellt. Im Idealfall ist dabei jedes einzelne *Datenfeld mit einem Trennzeichen* (z.B. Komma, Semikolon, Tabulator), mit einem zusätzlichen *Text-Einschlusszeichen* (z.B. Anführungszeichen bei freien Textinhalten in Buchungstext oder Adressen) sowie jeder vollständige Datensatz mit einem Satzende-Kennzeichen (häufig Sonderzeichen CR / LF) versehen.

```
*Test2.del - Editor
Datei Bearbeiten Format Ansicht Hilfe
"LIEFERNR";"RECHNUNGSB";"RECHNUNGSD";"FAELLIGEIT";"BELEGNR";"LIEFNAME";"ORT";"PLZ";"STRASSE";"VORWAHL";"TEU
"044550";384,96;"31.10.2003";"30.11.2003";234289;"A.B.C."; "Gütersloh";"33332";"Neuenkirchener Straße 55";"0
"471100";357,83;"16.04.2005";"16.05.2005";253470;"AKH FERRO";"Hamburg";"22769";"Haferweg 11";"040";1391735;
"471100";250,65;"09.04.2005";"09.05.2005";252555;"AKH FERRO";"Hamburg";"22769";"Haferweg 11";"040";1391735;
"471100";180,69;"27.04.2005";"27.05.2005";254139;"AKH FERRO";"Hamburg";"22769";"Haferweg 11";"040";1391735;
"022000";326,10;"16.04.2005";"29.06.2005";253124;"ALBAN";"Freiburg";"79104";"Weierhofstraße 2";"";327606;
"022000";172,43;"09.04.2005";"29.06.2005";252137;"ALBAN";"Freiburg";"79104";"Weierhofstraße 2";"";327606;
```

Abb. 8 Datei mit einem fehlerhaften Datenfeld ohne Einschlusszeichen und mit Feldtrenner als Inhalt

Probleme ergeben sich, wenn bei der Datenerzeugung eine unrichtige Anzahl erforderlicher Feldbegrenzer zu einzelnen Datensätzen berücksichtigt wurde. Hieraus folgen sinnentstellende Inhalte innerhalb betroffener Felder oder (bei Textdateien ohne Satzbegrenzer) der gesamten Datei. Soweit es sich um ein Einzelproblem handelt, welches nicht wiederholt und unkalkulierbar auftritt, hilft gegebenenfalls eine Durchsicht und manuelle Korrektur mittels geeignetem Editor.

	LIEFERNR	RECHNUNGSB	RECHNUNGSD	FAELLIGEIT	BELEGNR	LIEFNAME	ORT	PLZ	STRASSE	VORWAHL
1	044550	384,96	31.10.2003	30.11.2003	234289	A.B.C.	Gütersloh	33332	Neuenkirchener Straße 55	05241
2	471100	357,83	16.04.2005	16.05.2005	253470	AKH FERRO	Hamburg	22769	Haferweg 11	040
3	471100	250,65	09.04.2005	09.05.2005	252555	AKH FERRO	Hamburg	22769	Haferweg 11	040
4	471100	180,69	27.04.2005	27.05.2005	254139	AKH FERRO	Hamburg	22769	Haferweg 11	040
5	022000	326,10	16.04.2005	29.06.2005	253124	ALBAN	Freiburg	79104	Weierhofstraße 2	
6	022000	172,43	09.04.2005	29.06.2005	252137	ALBAN	Freiburg	79104	Weierhofstraße 2	

Abb. 9 Importierte fehlerhafte Datei mit Verschiebungen zu Feldinhalten nach der Fehlerposition

Eine häufige Ursache für das aufgeführte Problem besteht in der *fehlenden Abgrenzung von Feldern mit Textinhalten (Namen, Ortsangaben, Buchungstexten etc.) durch ein zusätzliches Text-einschlusskennzeichen*. Hierfür werden oft Anführungszeichen (s. Abb. 7) verwendet. Fehlt ein solches Kennzeichen generell oder gelegentlich und enthält der zugehörige Textinhalt zufällig einen Feldtrenner (z.B. Semikolon oder Komma), so wird dieses textliche Gestaltungsmerkmal seitens der Übernahmesoftware irrtümlich als Feldtrenner erkannt. Eine vergleichbare Ursache lässt sich auf die Verwendung von *Tabulatoren als Feldtrenner* zurückführen. Dieses ist besonders schwierig festzustellen, da der Text in vielen Editoren für den Bildschirm formatiert (und damit zutreffend) dargestellt wird, obwohl für eine zutreffende Ausrichtung - je nach Länge der vorhergehenden Spalte - ein oder mehrere Tabulatoren eingesetzt werden.

Soweit hiervon (bei genereller Abwesenheit von Text-Einschlusszeichen) nicht ausschließlich einzelne Datensätze betroffen sind, lohnt sich eine Korrektur kaum. In solchen Fällen ist es erfahrungsgemäß günstiger, eine neue Ausgangsdatei zu erzeugen.

Über die aufgeführten Probleme hinaus können sich in den hier behandelten Dateien (Text, CSV, Delimited) vergleichbare Unzulänglichkeiten (z.B. bei Fehlern in der Zeichentabelle) verbergen, wie in Textdateien mit fester Feld- und Satzlänge. Die bereitstehenden Lösungsoptionen sind identisch.

- **Druckausgaben als Text- und Ausgangsdateien**

Unabhängig von den bereits geschilderten Fehlerquellen in Textdateien treten bei druckaufbereiteten Ausgangsdaten (Listen, Formulare) *allgemeine Formatierungsprobleme* mit enthaltenen Deckblättern, Vorlaufinformationen, Drucksteuer- und sonstigen Sonderzeichen, Kopf-, Fuß- und Summenzeilen, Feldinhalten über mehrere Zeilen, Zeilenblöcke sowie unregelmäßige Zeilenabstände in den Vordergrund, die gelöst werden müssen, um den Daten wieder zu einer flachen (normalisierten und analysefähigen) Struktur zu verhelfen, bei der alle Informationen in einem einzelnen Datensatz gespeichert sind.

```

! Faelligkeit I bis..... 8 Tage !
! Faelligkeit II bis..... 30 Tage !
! OK _ PF: 3=Back 1 -38325 !
-----
^1BCI10;CL3;CS104;LI6^0D^0A16.11.22 OP - Faellikeits-Vorschau Debitoren
-----
^90000009 Hermann Hurtig Kreditlimit.....
^ Alles,,ufer & Co
^ Karlsruhe Versicherungslimit..
-----
||
^Letzte Mahnung... 01.08.22
^Mahnstufe ..... 4
-----
||
^ Offene Posten per 16.11.22
||
^ Off en ! Fa e l l i g
^ BK GB Ingesamt ! bis 8 Tage bis 30 Tage ueber 30 Tage
-----
! 01 ** ! 1.820,00 ! ! ! !
! 01 00 ! 1.820,00 ! ! ! !
-----
^16.11.22 OP - Faellikeits-Vorschau Debitoren 2
-----
^90000011 NEEF Kreditlimit.....
^ Karlsruhe Versicherungslimit..
-----
||
^ Offene Posten per 16.11.22
||
^ Off en ! Fa e l l i g
^ BK GB Ingesamt ! bis 8 Tage bis 30 Tage ueber 30 Tage
-----
! 01 ** ! 5.987,79 ! ! ! !
! 01 00 ! 5.987,79 ! ! ! !
-----

```

Abb. 10 Vorabkontrolle einer druckaufbereiteten Ausgangsdatei mit diversen Problembereichen in Texteditor

Entsprechenden Druckdateien erfordern eine Vorbehandlung, um sie in einen analysefähigen Zustand zu überführen. ACL und IDEA verfügen als tradierte Prüfsoftware über zugehörige Werkzeuge, welche bei *Druckdateien mit einem überschaubarem Maß an Unregelmäßigkeiten* ausreichende Lösungsmöglichkeiten eröffnen.

Wer davon unabhängig *nahezu alle Probleme*, die nach praktischer Erfahrung hiermit verbunden sind, lösen möchte und auf eine *einfache Handhabung* Wert legt, wird mit dem bereits bezeichneten „ImportWizard“ eine geeignete Lösungsoption finden. Diese beinhaltet auch die Berücksichtigung von Daten, die schwierigsten Formularcharakter aufweisen.

The screenshot shows the 'Import Designer' interface. The main window displays a text file with various fields. A 'Vorschau' (Preview) window is open, showing a table of data with columns: Row, BK, GB, Saldo, Bk_B, Bk_30, Ueber_30, Mahndat, Mahns, Name, Ort, KundenNr, Summe. The data rows show customer information for Hermann, NEEF, and NEEF.

The 'Import Wizard' dialog box is visible on the right, showing 'Importeinstellungen' (Import Settings) and 'Datei-Proprietäten' (File Properties). The 'Datei-Proprietäten' section shows 'Code Page' set to '437 OEM - United States (DOS)'. The 'Importeinstellungen' section shows 'Neuer Marker' (New Marker) and 'Neues Feld' (New Field) buttons.

Abb. 11 Kennzeichnung für die Analyse benötigten Felder und Inhalte mit dem ImportWizard (Importmaske)

Die in einer Übernahmeschablone ausgewählten Inhalte aus der druckaufbereiteten Datei können nachfolgend in jedes gängige und analysefähige Datenbankformat überführt werden. Die Übernahmeschablone kann gespeichert, wiederholt verwendet und automatisiert sowie mit beliebigen Filtern aufgerufen werden.

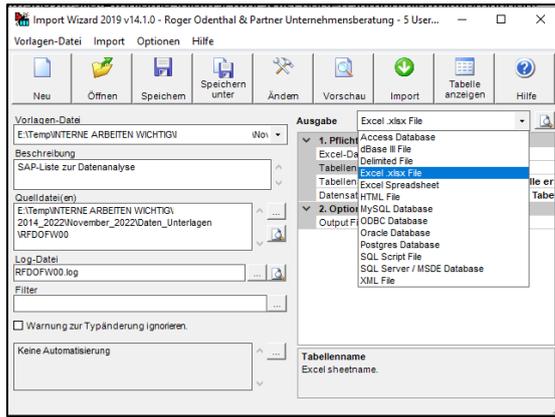


Abb. 12 Automatisierte Überführung der Druckdatei in ein gewünschtes Dateiformat (ImportWizard)

Name	Ort	KundenNr	BK	GB	Saldo	BH_8	BH_30	Ueber_30	Mahndatum	Mahnstufe	Summe
Hermann Hürtig Alleskäufer & Co	Karlsruhe	90000009	01	00	1.820,00	0,00	0,00	0,00	0,01.08.2022	00:00:00	4
NEEF	Karlsruhe	90000011	01	00	5.987,79	0,00	0,00	0,00			
griesbach gmbh	Karlsruhe	90000012	01	00	10.000,00	0,00	0,00	0,00			
Wicht Willi	Gartenstadt	90000014	01	00	711.670,84	0,00	0,00	0,00			
Wicht Willi	Gartenstadt	90000014	01	01	4.196.558,60	0,00	0,00	0,00			
Wicht Willi	Gartenstadt	90000014	01	02	115,00	0,00	0,00	0,00			
Wicht Willi	Gartenstadt	90000014	01	03	8.640,00	0,00	0,00	0,00			
Wachtel Walter	Heidelberg	90000016	01	00	39.400,00	0,00	0,00	0,00			
Wachtel Walter	Heidelberg	90000016	01	01	3.500,00	0,00	0,00	0,00			
Huber Handelshaus	Karlsruhe	90000017	01	00	12.000,00	0,00	0,00	0,00			
Wilma Wusel	Wurzelsbach	90000019	01	00	55.903,00	0,00	0,00	0,00			
khkh	sdalk	90000020	01	00	9.000,00	0,00	0,00	0,00			
Wieselfink GmbH & Co. KG Ersatzlieferant	Soforthause	90000021	01	00	-3,42	0,00	0,00	0,00			
Würmer Walter	Wurnsbach	90000024	01	00	1.931,39	0,00	0,00	0,00			
Manfred Maier	Karlsruhe	90000032	01	00	1.676,00	0,00	0,00	0,00			
Karl Wurster Wurstgroßhandlung	Fleischhaus	90000034	01	00	-177.700,00	0,00	0,00	0,00			
Dietmar Schucker	Pfingztal	90000039	01	00	-114.000,00	0,00	0,00	0,00			
Bauer-Sikora	Karlsruhe	90000042	01	02	6.255,65	0,00	0,00	0,00			
Mueller & Veith	Karlsruhe	90000044	01	00	1.210,00	0,00	0,00	0,00			
Mueller & Veith	Karlsruhe	90000044	01	01	11.500,00	0,00	0,00	0,00			
Mueller & Veith	Karlsruhe	90000044	01	05	15.000,00	0,00	0,00	0,00			
Stadtwerk Frankfurt an der Oder	Frankfurt	90000045	01	00	3.450,00	0,00	0,00	0,00			
Schrott & Partner Metallverwertung GmbH	Berlin	90000046	01	00	78.000,00	0,00	0,00	0,00			
Schrott & Partner Metallverwertung GmbH	Berlin	90000046	01	00	639.915,60	0,00	0,00	0,00			SUMMEN
Schrott & Partner Metallverwertung GmbH	Berlin	90000046	01	01	4.211.558,60	0,00	0,00	0,00			SUMMEN
Schrott & Partner Metallverwertung GmbH	Berlin	90000046	01	02	6.255,65	0,00	0,00	0,00			SUMMEN
Schrott & Partner Metallverwertung GmbH	Berlin	90000046	01	03	8.640,00	0,00	0,00	0,00			SUMMEN
Schrott & Partner Metallverwertung GmbH	Berlin	90000046	01	05	15.000,00	0,00	0,00	0,00			SUMMEN

Abb. 13 Ergebnis der Überführung der Druckdatei in ein analysefähiges Excelformat

Ein weiteres Problem druckaufbereiteter Daten sind mögliche Verschiebungen innerhalb einzelner Spaltenbereiche, die zum Beispiel durch Drucksteuerzeichen oder Formatierungen entstehen können.

LIEFERNR	RECHNUNGSBRECHUNGSD	FAELLIGEIT	BELEGNRLIEFNAME	ORT	PLZ	STRASSE
044550	384,9631.10.2003	30.11.2003	234289A.B.C.	Gütersloh	33332	Neuenkirchener Straße
471100	357,8316.04.2005	16.05.2005	253470AKH FERRO	Hamburg	22769	Haferweg 11
4711	250,6509.04.2005	09.05.2005	252555 AKH FERRO	Hamburg	22769	Haferweg 11
471100	180,6927.04.2005	27.05.2005	254139AKH FERRO	Hamburg	22769	Haferweg 11
022000	326,1016.04.2005	29.06.2005	253124ALBAN	Freiburg	79104	Weierhofstraße 2

Abb. 14 Typische Verschiebungen und Fehler in schlecht aufbereiteten Druckdateien

In diesem Fall hilft es in der Regel, die für einzelne Spaltenfelder definierte Spaltenbreite innerhalb des ImportWizard so weit zu fassen, dass sie die entsprechenden Verschiebungen zusätzlich berücksichtigt.

LIEFERNR	RECHNUNGSBRECHUNGSD	FAELLIGEIT	BELEGNRLIEFNAME	ORT	PLZ	STRASSE
044550	384,9631.10.2003	30.11.2003	234289A.B.C.	Gütersloh	33332	Neuenkirchener Straße 55
471100	357,8316.04.2005	16.05.2005	253470AKH FERRO	Hamburg	22769	Haferweg 11
4711	250,65 09.04.2005	09.05.2005	252555 AKH FERRO	Hamburg	22769	Haferweg 11
471100	180,6927.04.2005	27.05.2005	254139AKH FERRO	Hamburg	22769	Haferweg 11
022000	326,1016.04.2005	29.06.2005	253124ALBAN	Freiburg	79104	Weierhofstraße 2
022000	172,4309.04.2005	29.06.2005	252137ALBAN	Freiburg	79104	Weierhofstraße 2

Abb. 15 Berücksichtigung der aufgeführten Fehler bei der Felddefinition in ImportWizard

In dem Übernahmeergebnis werden die Daten anschließend zutreffend und einheitlich dargestellt.

LieferantenNr	Betrag	Rechnungsdatum	Faelligkeit	BelegNr	Name	Ort	PLZ	Strasse	Vorwahl	Telefon
044550	384,96	31.10.2003	30.11.2003	234289	A.B.C.	Gütersloh	33332	Neuenkirche 05241		841466
471100	357,83	16.04.2005	16.05.2005	253470	AKH FERRO	Hamburg	22769	Haferweg 11 040		1391735
4711	250,65	09.04.2005	09.05.2005	252555	AKH FERRO	Hamburg	22769	Haferweg 11 040		1391735
471100	180,69	27.04.2005	27.05.2002	254139	AKH FERRO	Hamburg	22769	Haferweg 11 040		1391735
022000	326,1	16.04.2005	29.06.2005	253124	ALBAN	Freiburg	79104	Weierhofst		327606
022000	172,43	09.04.2005	29.06.2005	252137	ALBAN	Freiburg	79104	Weierhofst		327606

Abb. 16 Ergebnis der Datentransformation mit automatischer Fehlerkorrektur durch den ImportWizard

Der aufgeführte Lösungsweg führt nicht zum Ziel, wenn die Verschiebungen in den Spaltenbereich anderer Felder hineinreichen. Auch in solchen Fällen ist eine Datenübernahme möglich, jedoch müssen hierbei abweichende Importstrategien verfolgt werden.

Marker	10	20	30	40	50	60	70	80	90	100
1	LIEFERNR	RECHNUNGSRECHNUNGSD	FAELIGEIT	BELEGNRLIEFNAME	ORT					
2	044550	384,9631.10.2003	30.11.2003	234289A.B.C.	Gütersloh					
3	471100	357,8316.04.2005	16.05.2005	253470AKH FERRO	Hamburg					
4	4711	250,65 09. 04.2005	09.05.2005	252555 AKH FERRO	Hamburg					
5	471100	180,6927.04.2005	27. 05.2 005	254139AKH FERRO	Hamburg					
6	022000	326,1016.04.2005	29.06.2005	253124ALBAN	Freiburg					
7	022000	172,4309.04.2005	29.06.2005	252137ALBAN	Freiburg					
8	022000	24,4030.03.2005	30.05.2005	250889ALBAN	Freiburg					
9	022000	9,7881.03.2005	30.05.2005	251447ALBAN	Freiburg					

Abb. 17 Beispiel für Verschiebungen und Fehler, die in weitere Spalten hineinreichen

Hierfür stehen bei Nutzung des ImportWizard unterschiedliche Optionen zur Verfügung:

- Verwendung unterschiedlicher Marker (Selektoren)

Die betroffenen Datensätze können mit *getrennten Markern ausgewählt und gesondert* übernommen werden. Der Import Wizard unterstützt ein entsprechendes Vorgehen durch die Berücksichtigung einer *beliebigen Anzahl von Selektoren* und diesen zugeordneten Feldern für unterschiedlichste Ausprägungen bei den Datensätzen. Felder mit zueinanderpassenden Inhalten können dann in einem zweiten Schritt (nach dem Import) innerhalb von Prüfsoftware zusammengefügt werden. Das aufgeführte Vorgehen eignet sich dann, wenn die Verschiebungen selbst eine gewisse Regelmäßigkeit aufweisen. Ansonsten wäre es zu aufwendig.

- Import ganzer Datensätze in einem Textformat mit anschließender Felddifferenzierung

Es handelt sich ebenfalls um ein *zweistufiges Verfahren*, welches Datensätze zunächst nach Prüfsoftware importiert, um dort mit Hilfe des Gleichungseditors einzelne (Rechen-) Felder anzulegen, die durch den Einsatz von Formeln und Funktionen mit den zutreffenden Inhalten versorgt werden. Sie entspricht (auf etwas umständlicherer Weise) dem nachfolgend geschilderten Vorgehen.

- Zusätzliche überschlägige Felddefinitionen und Anlage von „Formelfeldern“ (ImportWizard)

Innerhalb des ImportWizard werden hier alle zu importierenden und unproblematischen Spalten „normal“ definiert. In vorstehendem Beispiel ist dieses für nahezu alle Spalten (bis auf Rechnungsbetrag und Rechnungsdatum) möglich. Die „problematischen“ Spalten mit ineinanderfließenden Inhalten können zunächst als ein zusammenhängendes Textfeld definiert werden:

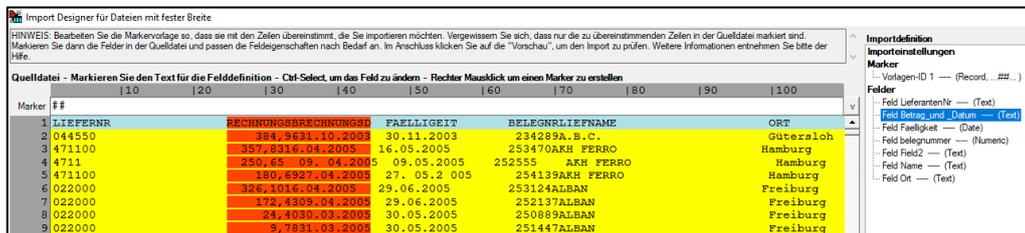


Abb. 18 Beispiel für überschlägige Felddefinitionen bei ineinanderfließenden Verschiebungen

Anschließend werden (für das vorstehende Beispiel) zwei neue Felder „Rechnungsbetrag“ und „Rechnungsdatum“ als Formelfelder definiert. Der ImportWizard unterstützt *bereits bei der Datenübernahme sämtliche Formeln und Funktionen*, die VBA (als Programmiersprache) bereitstellt:

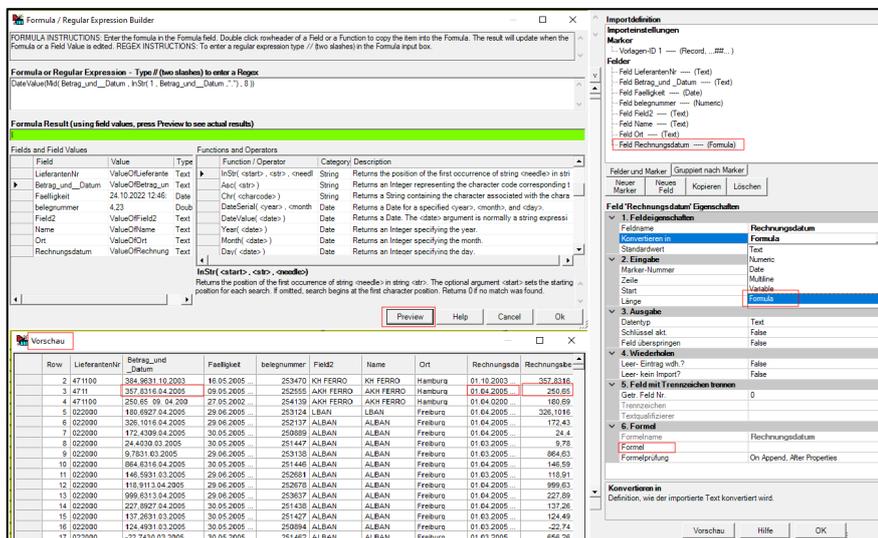


Abb. 19 Beispiel für die zutreffende Interpretation von Importfeldern bei Formeleinsatz (ImportWizard)

Die einem Selektor mit „Formeln“ zugeordneten Felder werden während des anschließenden Importprozesses, analog zu allen weiteren Feldern, in das gewünschte Format überführt.

LieferantenNr	Betrag_und_Datum	Faelligkeit	belegnummer	Name	Ort	Rechnungsbetrag	Rechnungsdatum
044550	384,9631.10.2003	30.11.2003	234289	A.B.C.	Gütersloh	384,96	31.10.2003
471100	357,8316.04.2005	16.05.2005	253470	KH FERRO	Hamburg	357,83	16.04.2005
4711	250,65 09. 04.2005	09.05.2005	252555	AKH FERRO	Hamburg	250,65	09.04.2005
471100	180,6927.04.2005	27.05.2002	254139	AKH FERRO	Hamburg	180,69	27.04.2005
022000	326,1016.04.2005	29.06.2005	253124	LBAN	Freiburg	326,10	16.04.2005
022000	172,4309.04.2005	29.06.2005	252137	ALBAN	Freiburg	172,43	09.04.2005
022000	24,4030.03.2005	30.05.2005	250889	ALBAN	Freiburg	24,40	30.03.2005
022000	9,7831.03.2005	30.05.2005	251447	ALBAN	Freiburg	9,78	31.03.2005
022000	864,6316.04.2005	29.06.2005	253138	ALBAN	Freiburg	864,63	16.04.2005

Abb. 20 Ergebnis des Importprozesses mit Formeleinsatz (ImportWizard)

Die von dem ImportWizard bereitgestellte Übernahmemaske stellt die aufgeführte Transformation auch bei mehrfacher und automatisierter Anwendung sicher.

Mit den hier aufgeführten Techniken lassen sich bereits eine Vielzahl von Unzulänglichkeiten in schwierigen Ausgangsdateien beheben. Mit etwas Übung eröffnen sie auch in kritischen Fällen noch Lösungen, die sich mit überschaubarem Aufwand umsetzen lassen.

- **Ausgangsdaten in einem PDF-Format**

Ausnahmslos alle der bisher aufgeführten Probleme können in vergleichbarer Form innerhalb von Ausgangsdaten auftreten, die in einem der zahlreichen „PDF-Formate“ für Analysezwecke bereitgestellt werden. Gleiches gilt auch für hierzu beschriebene Lösungsempfehlungen, die in identischer Form auf PDF-Dateien übertragen werden können. Hinzu gesellen sich dann einige Schwierigkeiten, die exklusiv dem PDF-Format zugerechnet werden können, und auf die wir nachfolgend eingehen:

- PDF ist nicht PDF

Bei PDF (Portable Document Format) handelt es sich um eine vektorbasierte Seitenbeschreibungssprache, welche frei skalierbare Darstellungen von Dokumenten (z.B. Drucklisten, Formularen, Ausdrucken etc.) erlaubt. Unabhängig von dem allgemein verwendeten Suffix „PDF“ werden die hierin enthaltenen Daten für unterschiedlichste Zwecke in unterschiedlichsten Varianten (PDF/A, PDF/UA, PDF/X, PDF/VT und PDF/E) angeboten. Nicht alle Formate werden von den Importfunktionen der verwendeten Prüfsoftware beherrscht und können zutreffend interpretiert werden.

Selbst bei Verwendung eines gängigen (genormten) Formats ist nicht sichergestellt, dass sich alle erzeugenden (buchhaltungsrelevanten oder rechnungslegende) Programme an die Vorgaben der einzelnen PDF-Normen halten. Die Folge sind zahlreiche Fehlerquellen, welche eine technisch und inhaltlich unproblematische Übernahme von Ausgangsdaten für Prüfungszwecke behindern¹.

- Der Bildschirm des Import-Programms bleibt leer

Ein unangenehmer Vorgang, bei welchem die PDF-Datei offensichtlich Daten enthält, die zudem in einem PDF-Reader zutreffend dargestellt werden, während sie von den Import-Funktionen der Prüfsoftware (ACL, IDEA bis hin zu ImportWizard) ignoriert werden. Der Bildschirm bleibt leer oder enthält lediglich einige sinnentstellende Zeichen:

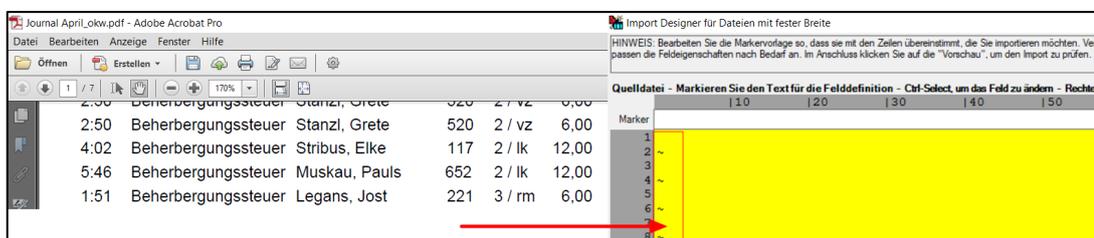


Abb. 21 PDF-Dokument führt zu leerem Bildschirm bei Importfunktionen von Prüfsoftware

Die möglichen Ursachen sind vielfältig und reichen von ungültigen Schriften über falsche Kompression von Bildformaten, Transparenzeigenschaften bis zu Verschlüsselungen.

Die aufgeführte Unzulänglichkeit kann häufig mit einem sachgerechten PDF-Tool (z.B. Acrobat Pro) gelöst werden, welches in der Lage ist, die PDF-Angaben in einem formatierten RTF-, Word- oder

¹ Eine Erläuterung zu häufigen Fehlerquellen wird über die folgende Seite angeboten:
https://www.slub-dresden.de/fileadmin/groups/slubsite/Service/PDF_Service/Eräuterung_häufiger_Fehlermeldungen.pdf

Excel-Format auszugeben. Innerhalb von Word oder Excel sollte das Ergebnis anschließend als formatierte Bildschirmausgabe erscheinen.

Hotel "Ringstrasse", Kramplin							Seite 8 von 122	
Journal nach Warengruppe vom 01.04.2022 bis 30.04.2022							Druck: 08.06.2022 - 14:44 / Josef Ströb	
Zeit	Buchungstext	Gastname	Zimmer #	Kasse / User	Debit brutto	Debit netto	Kredit	
01.04.2022								
920 - Kulturabgabe								
2:50	Kulturabgabe	Stanzl, Grete	520	2 / vz	6,00	6,00		
2:50	Kulturabgabe	Stanzl, Grete	520	2 / vz	6,00	6,00		
4:02	Kulturabgabe	Stribus, Elke	117	2 / lk	12,00	12,00		
5:46	Kulturabgabe	Muskau, Pauls	652	2 / lk	12,00	12,00		
1:51	Kulturabgabe	Legans, Jost	221	3 / rm	6,00	6,00		
Subtotal Kulturabgabe					01.04.2022	42,00	42,00	0,00
Total 01.04.2022						42,00	42,00	0,00

Abb. 22 Word-Dokument nach der Transformation aus einem PDF-Tool

Diese kann anschließend in Word oder Excel mit der Funktion „Speichern unter“ wiederum in ein (nun interpretierbares) PDF-Format umgewandelt werden, welche seitens Prüfsoftware oder dem ImportWizard nun in einer lesbaren Form zahlreiche Möglichkeiten eröffnet, um diese in eine analysefähige Datei umzuwandeln.

Abb. 23 Analysefähige PDF-Daten nach der Transformation in der Software ImportWizard (Vorschau)

Das aufgeführte Vorgehen erfordert keine besondere technische Expertise und führt nach praktischer Erfahrung oft zu einem brauchbaren Ergebnis.

- Die PDF-Datei weist einen Zeilenversatz auf, welcher den Import behindert

Eine solche Situation kann auftreten, wenn die PDF-Erzeugung aus dem Ausgangssystem mit einem zu geringen Zeilen- oder Zeichenbreite erfolgte.

Abb. 24 Zur Analyse überlassene PDF-Ausgangsdatei mit Zeilenversatz und Lösungsmöglichkeit

Der ImportWizard ermöglicht eine freie Einstellung des Zeichenabstandes innerhalb des PDF-Dokuments. Bei Eingabe einer „0“ erfolgen die Kalkulation und Ausrichtung automatisch.

Hotel "Ringstrasse", Kramplin							Seite 8 von 122	
Journal nach Warengruppe vom 01.04.2022 bis 30.04.2022							Druck: 08.06.2022 - 14:44 / Josef Ströb	
Zeit	Buchungstext	Gastname	Zimmer #	Kasse / User	Debit brutto	Debit netto	Kredit	
01.04.2022								
920 - Kulturabgabe								
2:50	Kulturabgabe	Stanzl, Grete	520	2 / vz	6,00	6,00		
2:50	Kulturabgabe	Stanzl, Grete	520	2 / vz	6,00	6,00		
4:02	Kulturabgabe	Stribus, Elke	117	2 / lk	12,00	12,00		
5:46	Kulturabgabe	Muskau, Pauls	652	2 / lk	12,00	12,00		
1:51	Kulturabgabe	Legans, Jost	221	3 / rm	6,00	6,00		

Abb. 25 Beseitigter Zeilenversatz in einem PDF-Dokument bei neuem Zeichenabstand (ImportWizard)

- Die PDF-Datei enthält schlecht eingescannte Dokumente, seltene Schrifttypen oder Grafiken

Ein weiteres Problem druckaufbereiteter und hier insbesondere von PDF-Daten sind unterschiedliche Schrifttypen oder eingebettete Grafiken, welche in die Druckausgabe integriert wurden. Diese Probleme behandelt der ImportWizard dadurch, dass er die eingegebene Druckdatei auf *eine einheitliche Zeichenbasis transformiert und Grafiken nicht berücksichtigt*.

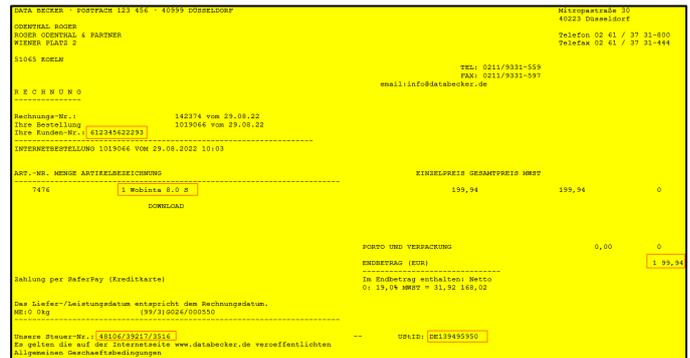
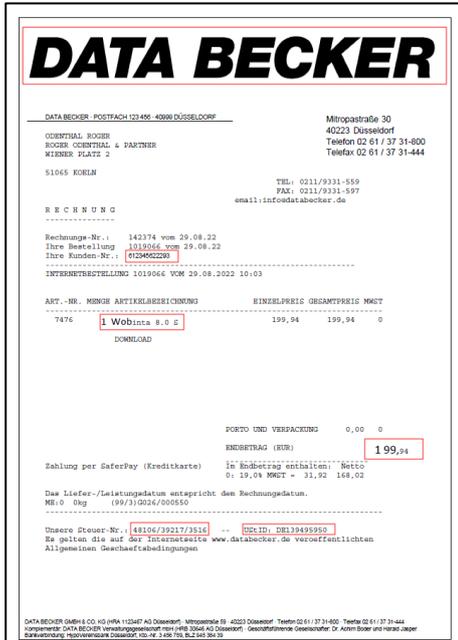


Abb. 26 PDF-Ausgangsdatei mit Überschriften, Lücken, verschiedenen Schrifttypen in ImportWizard

Damit lösen sich einige Problembereiche ohne aufwendige Aktivitäten. Nun gilt es noch, die auf den ersten Blick völlig unstrukturierten Daten zu übernehmen und hierbei auch Formulare mit vergleichbaren Inhalten aber unterschiedlichem Aufbau (z.B. bei Kontenauszügen) zu berücksichtigen. Hier offeriert der Import Wizard eine *Lösung auf der Basis „regulärer Ausdrücke“*, die zur *Gestaltung der Marker (Selektoren) verwendet werden können*. Sie ermöglichen die Erstellung analysefähiger Daten aus weitgehend unstrukturiertem Text, wie das folgende Beispiel zeigt:

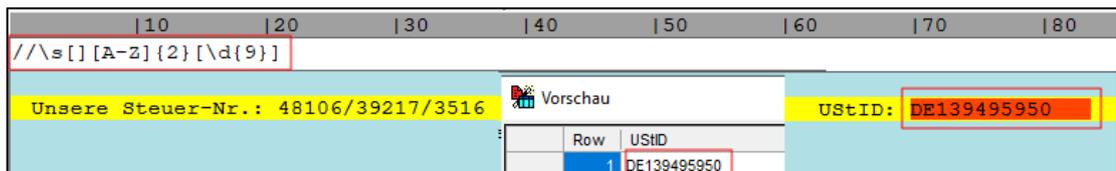


Abb. 27 Marker und Feld aus einer beliebigen Zeile mit Leerstelle, zwei Großbuchstaben und 9 Ziffern

Wir gehen in einem gesonderten Kapitel auf die Möglichkeiten regulärer Ausdrücke ein und beschreiben dort die Struktur anhand mehrerer Beispiele.

PDF-Daten sind die für Revisorinnen und Revisoren am wenigsten kalkulierbare Datengrundlage für prüferische Datenanalysen. Im Hinblick auf die zahlreichen technischen Aspekte sollten sie lediglich dann verwendet werden, wenn keine alternative und leichter handhabbare Datenquelle zu Verfügung steht.

- **Excel Dateien, die nicht in einem Datenbankformat vorliegen**

Excel-Dateien gelten üblicherweise als unproblematisch. Dieses gilt jedoch lediglich so lange, wie die Daten in einem „Datenbankformat mit flacher Struktur“ angeboten werden. Hier folgen auf die Feldbezeichnungen in der ersten Zeile einer Tabelle ausschließlich noch Daten und keine Umbrüche, Zwischensummen oder Querbezüge. Tatsächlich werden Excel-Daten in der Prüfungspraxis jedoch in unterschiedlichster Form bis hin zu komplexen druckaufbereiteten Formularen mit Grafiken usw. angeliefert, die zunächst aufwendig und fehlerfrei eliminiert werden müssen.

Abb. 28 Aus einem SAP-Report erzeugte Excel-Datei mit heterogenen Satzstrukturen

Prüfsoftware stellt für entsprechende Vorbereitungen und Bereinigungen keine gesonderte Funktionalität bereit. Selbst Excel bietet keine vernünftige Lösung an, so dass auch an dieser Stelle der ImportWizard das Werkzeug der Wahl ist. Schließlich bietet er für Excel-Daten (bezogen auf alle Spalten, Zeilen und Zellen) die gleichen Optionen zur automatisierten Vorbehandlung an, die auch für Ausgangsdaten in Text-, Druck- und PDF-Formaten bereitstehen. Hierdurch lassen sich nahezu alle aufgeführten Probleme, die im Zusammenhang mit Excel-Daten auftauchen können, in ähnlich eleganter Form lösen.

Abb. 29 Übernahme und Vorbehandlung einer Excel-Datei mit Hilfe des ImportWizard

Mit dem aufgeführten Vorgehen gehören unzuverlässige Formatierungen oder aufwendige Vorarbeiten bei unterschiedlich Satzstrukturen sowie bei der Zuordnung von Kopf- oder Summenpositionen zu Detailzeilen in Excel-Dateien der Vergangenheit an. Das Ergebnis wird immer in eine analysefähige flache Datenbankstruktur überführt.

• Fehler bei der Übernahme von GoBD-Daten

Auf mögliche Probleme im Zusammenhang mit bereitgestellten GoBD-Dateien haben wir bereits in den Ausführungen zu geeigneten Werkzeugen hingewiesen. Hier bieten sowohl der GoBD-Importer als auch der ImportWizard wertvolle Unterstützung, wie die folgenden Detailbetrachtungen zu möglichen Fehlerquellen demonstrieren:

- Erste Sichtkontrolle des GoBD-Datensets

Abb. 30 GoBD-Importer mit Auswahl zu importierender Daten und Fehlerprotokoll nach dem Import

Der GoBD-Importer vermittelt einen Überblick über enthaltene Daten und erzeugt ein Fehlerprotokoll, wenn der Datenimport nicht einwandfrei funktionierte. Zusätzlich werden die Inhalte der einzelnen Tabellen in ein gewünschtes Format (Excel, Access) überführt und können dort verfolgt oder für Analyse Zwecke eingesetzt werden.

Ktonr	Text	ZusFkt	HkTyp	HFNr	FktErg	F2Proz	F2Kto1	F2Kto2	IFktBeschr	AnspFkt
135	EDV-Software	0	0	0	0	0,00	0	0		0
140	Lizenzen an gewerblichen Schutzrechten	0	0	0	0	0,00	0	0		0
520	PKW	0	0	0	0	0,00	0	0		0
650	Büroeinrichtung	0	0	0	0	0,00	0	0		0
670	Geringwertige Wirtschaftsgüter	0	0	0	0	0,00	0	0		0
675	Wirtschaftsgüter Sammelposten	0	0	0	0	0,00	0	0		0
690	Sonstige Betriebs- u. Gesch.ausstattung	0	0	0	0	0,00	0	0		0
980	Genossenschaftsanteile z.lfr.Verbleib	0	0	0	0	0,00	0	0		0
1.200	Forderungen aus Lieferungen u. Leistung	0	0	0	0	0,00	0	0		0
1.350	Kauttionen	0	0	0	0	0,00	0	0		0
1.370	Durchlaufende Posten	0	0	0	0	0,00	0	0		0
1.371	Unklare Posten	0	0	0	0	0,00	0	0		0
1.401	Abziehbare Vorsteuer 7%	0	0	0	0	0,00	0	0		0

Abb. 31 Inhalt einer einzelnen Datei eines GoBD-Datensets (GoBD-Importer)

Erste mögliche Fehler sollten hiernach bereits identifizierbar sein. Sie erfordern ggf. eine detaillierte Analyse, die sich nun anschließen kann:

- Nach Auswahl der „Index.XML“-Datei werden zugehörige Buchhaltungsdaten nicht gefunden
- Die grundsätzlichen Angaben zum Aufbau der Index.XML-Datei und der Datendateien finden sich in der begleitenden GoBD-Schema-Datei (z.B. GDPDU-01-08-2002.dtd).

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Versions available:
1.1 (August-01-2002)
-->

<!-- Start Simple Types -->
<!-- Supplementary Vocabulary -->
<ELEMENT Version (#PCDATA)>
<ELEMENT Location (#PCDATA)>
<ELEMENT Comment (#PCDATA)>
<ELEMENT Length (#PCDATA)>
<ELEMENT References (#PCDATA)>
<ELEMENT From (#PCDATA)>
<ELEMENT To (#PCDATA)>

<!-- Specifying a maximum length for a VariableLength column can
reduce a VariableLength tables' import time. If MaxLength
is not specified then we parse URL to determine the MaxLength
for each column.

* Only applies to VariableLength tables. -->
<ELEMENT MaxLength (#PCDATA)>

<!-- Specifies which character (if any) encapsulates a
VariableLength AlphaNumeric column.

Doublequote is the default TextEncapsulator "

* Only applies to VariableLength tables. (Optional) -->
<ELEMENT TextEncapsulator (#PCDATA)>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE DataSet SYSTEM "gdpdu-01-08-2002.dtd">
<DataSet>
  <Version>1.0</Version>
  <DataSupplier>
    <Name>DATEV eG</Name>
    <Location>90329 Nürnberg</Location>
    <Comment>Kanzlei-Rechnungswesen pro 4.0A / ASCII-Export</Comment>
  </DataSupplier>
  <Media>
    <Name />
    <Table>
      <URL>Sachkontenstamm.csv</URL>
      <Name>Sachkontenplan</Name>
      <Description>Kontenplan der Sachkonten</Description>
      <DecimalSymbol>,</DecimalSymbol>
      <DigitGroupingSymbol />
      <VariableLength>
        <VariableColumn>
          <Name>Ktonr</Name>
          <Description>Kontonummer des Kontos</Description>
          <Numeric />
        </VariableColumn>
      </VariableLength>
    </Table>
  </Media>
</DataSet>
```

Abb. 32 Blick mit einem Editor in die Schema-Datei (links) und die Index.xml (rechts)

In den einzelnen Segmenten der Index.xml erhält die jeweils verwendete Importfunktion Informationen zum Speicherort enthaltener Datendateien, zu deren Format, zu verwendeten Feldtrennern, Dezimalzeichen sowie zu den enthaltenen Feldern. Stimmen die Angaben in der Sektion „DataSupplier“ nicht mit dem tatsächlichen Speicherort überein, so ist der aufgeführte Fehler die Folge. Mit einem Texteditor können in der Index.xml sowohl grundsätzliche Angaben (Media, Name, Table) als auch die Bezeichnungen zu enthaltenen Einzeldateien (URL) angepasst werden. Wenn eine Ausgangsdatei Bestandteil des überlassenen GoBD-Sets ist und nicht lediglich als Beschreibung in der Index.xml auftaucht (was eine zusätzliche Kontrolle erfordert), sollte der Import hiernach grundsätzlich funktionieren.

- Fehlerhafte Feldinhalte oder Formate nach dem Import der GoBD-Daten

Eine verbreitete Fehlerquelle beruht auf Divergenzen zwischen der inhaltlichen Beschreibung (Satzbett) der Ausgangsdaten innerhalb der Index.xml und deren tatsächlichem Aufbau. Für eine hierauf gerichtete Kontrolle empfiehlt es sich zunächst die „Index.xml“ in den ImportWizard zu laden, um einen Überblick über deren Gesamtstruktur zu erhalten:

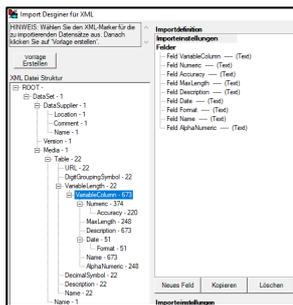
Abb. 33 Darstellung der Struktur der Index.xml und der erzeugten Vorlage (ImportWizard)

Die Erstellung von Vorlagen folgt der XML-Struktur. Auf dem Knoten „Table“ werden alle Felder zur Beschreibung der Ausgangsdaten geliefert und mit Inhalten ausgegeben:

URL	Accuracy	MaxLength	Description	Date	Format	Name	DecimalSymbol
Sachkontenstamm.csv	2	40	Anlagentypfunktion			AnspFkt	,
DebitorenKreditorenstammdaten.csv		54	Bankverbindung 5 gueltig bis		DD.MM.YYYY	Bankv5_Gueltig_bis	,
Kontobuchungen.csv	2	9	Gewicht		DD.MM.YYYY	Gewicht	,
Buchungssatzprotokoll.csv	2	9	Gewicht		DD.MM.YYYY	Gewicht	,
Bereichsuebersicht.csv		40	Zweck der Rechnungslegung			Rechnlegungszweck	,
Mvz.csv	2	40	Monatsverkehrszahlen Monat 12 Haben		DD.MM.YYYY	MVZ_12_H	,
Mandantendaten.csv		12	Festschreibung erfolgt (1=alles festgeschrieben, 0=nicht alles festge:		DD.MM.YYYY	Festschreibung	,
Kontennachweis.csv	2	100	Hauptspalte NK 2 (Ausgabewert)			KNW_Hauptspalte	,
Buchungstapelliste.csv		11	Anzahl Belege im Stapel		DD.MM.YYYY	Anzahl_Belege	,
InventarStamm.csv	2	60	Investitionsabzugsbetrags-ID		DD.MM.YYYY	InvestID	,

Abb. 34 Hinweise zu enthaltenen Tabellen in der „Index.xml“ (ImportWizard)

Die Positionierung des Cursors auf den Knoten „VariableColumn“ gefolgt von dem Schalter „Vorlage erstellen“ führt zu einer Übersicht über die einzelnen Felder enthaltener Tabellen:



URL	Accuracy	MaxLength	Description	Name
Sachkontenstamm.csv			Kontonummer des Kontos	Ktonr
		40	Beschreibung	Text
			ZusFkt	ZusFkt
			Hauptfunktionstyp	HkTyp
			HFNr	HFNr
			Funktionsergaenzung	FktErg
	2		Faktor2_Prozent	F2Proz
			KtoNr1_Faktor2	F2Kto1
			KtoNr2_Faktor2	F2Kto2
		40	Funktionsbezeichnung	IFktBeschr
			Anlagentypfunktion	AnspFkt

Abb. 35 Darstellung der Struktur der Index.xml und der erzeugten Vorlage (ImportWizard)

Das Feld URL (Tabellenbezeichnung) sollte mit dem zugehörigen XML-Pfad ergänzend in die Feldübersicht der Importeinstellungen übernommen werden.

Mit den nun vorliegenden Informationen können anschließend die konkreten Inhalte der in dem GoBD-Set enthaltenen Datendateien abgeglichen werden. In der Regel wird es sich um Textdateien mit variablen Feldlängen und Satztrenner (CSV, Delimited) handeln, die ebenfalls mittels des ImportWizard eingeladen werden:

Field1	Field2	Field3	Field4	Field5	Field6	Field7	Field8	Field9	Field10	Field11
135	EDV-Software	0	0	0	0	0,00	0	0		0
140	Lizenzen an gewerblichen Schutzrechten	0	0	0	0	0,00	0	0		0
520	PKW	0	0	0	0	0,00	0	0		0
650	Büroeinrichtung	0	0	0	0	0,00	0	0		0
670	Geringwertige Wirtschaftsgüter	0	0	0	0	0,00	0	0		0
675	Wirtschaftsgüter Sammelposten	0	0	0	0	0,00	0	0		0
690	Sonstige Betriebs-u.Gesch.ausstattung	0	0	0	0	0,00	0	0		0
980	Genossenschaftsanteile z.lfr.Verbleib	0	0	0	0	0,00	0	0		0

Abb. 36 Kontrolle der Datendatei eines GoBD-Sets - Anzahl Felder und Inhalte (ImportWizard)

Ergeben sich Differenzen bei der Anzahl von Feldern, Feldinhalten oder -typen, so können (bei Einzelfällen) Korrekturen mittels eines Editors vorgenommen werden. Ansonsten ist es sinnvoll, eine neue Ausgangsdatei anzufordern.

Hinsichtlich inhaltlicher Fehler in den bereitgestellten GoBD-Daten kann an dieser Stelle auf alle bereits aufgeführten Problembereiche zu Textdateien verwiesen werden, die bereits vorgestellt wurden. Die Lösungsmöglichkeiten und Werkzeuge sind ebenfalls identisch.

3 Reguläre Ausdrücke (Regular Expressions) bei schwierigen Importen

„Reguläre Ausdrücke“ bezeichnen eine formale Beschreibungssprache (mit vorgegebenen syntaktischen Regeln) zur Identifizierung von Zeichenmustern innerhalb beliebiger Zeichenketten. In dieser abstrakten Form ist sie Gegenstand der theoretischen Informatik und unterstützt insbesondere die Softwareentwicklung bei der Erstellung von Prozeduren, die in zahlreichen Softwarekategorien (Texteditoren, Prüfsoftware, Analyseprogrammen) Verwendung finden. Anwendern begegnet sie hier in abgewandelter Form, z.B. in Suchmenüs oder in den Textfunktionen von Gleichungseditoren. Eine Verwendung in „Rohform“, z.B. zur Selektion komplexer Textmuster innerhalb unstrukturierter Ausgangsdateien, ist insbesondere für Auditzwecke unter zwei Voraussetzungen nützlich. Zunächst muss die eingesetzte Software eine solche Anwendung bereits während des Imports (und nicht nur zur Behandlung einer bereits übernommenen Datei) unterstützen und die jeweiligen Anwender sollten

eine belastbare Vorstellung von den syntaktischen Grundlagen der Beschreibungssprache haben. Trifft beides zu, dann lassen sich nahezu beliebige Daten automatisiert identifizieren und einem Datenfeld zuordnen. Es handelt sich um ein komplexes, aber für die Prüfung wichtiges Thema. Die folgenden Beispiele geben daher (ohne Anspruch auf Vollständigkeit) einen kurzen Einblick in das Thema:

- **Unterstützung in Prüfsoftware**

Der ImportWizard berücksichtigt reguläre Ausdrücke bereits bei der Gestaltung seiner Marker (Selektoren) und darüber hinaus als Bestandteil des inhärenten Gleichungseditors innerhalb von Formelfeldern sowie von Selektionsfiltern. Alle aufgeführten Anwendungen liegen vor der Transformation von Ausgangsdateien:

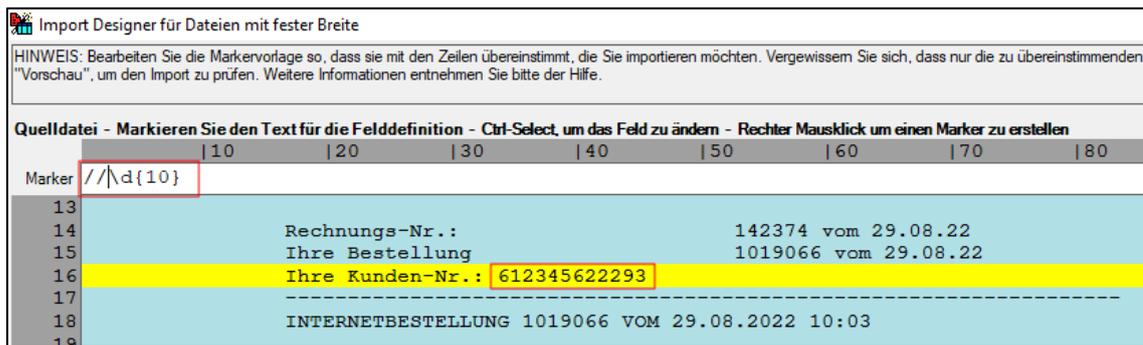


Abb. 37 Anwendung regulärer Ausdrücke bei komplexen Suchmustern innerhalb des ImportWizard

ACL verwendet für Selektoren (vor der Transformation) mit variablen Suchmustern die Funktion „MAP()“ und stützt sich hierbei, ähnlich wie der ReportReader von IDEA, weitgehend auf Standardvorgaben, die keine Entsprechung zu „regulären Ausdrücken“ haben. Erst nach dem Import (Transformation) können bei ACL innerhalb von Rechen- und Filterformeln zu Suchmustern die Funktion REGEXREPLACE() und bei IDEA die Funktion REGEXPR() angewendet werden. Für die Anwendung bei schwierigen Dateien ist der nachgeordnete Einsatz regulärer Ausdrücke allerdings wenig hilfreich. Zu hier jeweils zugeordnete Hilfe vermittelt jedoch einen Eindruck zur Leistungsfähigkeit dieses bisher wenig beachteten Werkzeugs.

- **Grundlegende Notation bei regulären Ausdrücken**

Bei regulären Ausdrücken handelt es sich um eine formale Beschreibungssprache (Beschreibungslogik) für Textmuster in beliebigen, häufig unstrukturierten Zeichenketten (Dokumenten). Die Beschreibung erfolgt mittels normierter syntaktischer Regeln. Die aufgeführten Beschreibungsregeln werden in erster Linie zur Entwicklung anwenderbezogener Prozeduren in Applikationen (z.B. Suchen, Ersetzen, Filtern) oder zur Ausstattung von Programmiersprachen genutzt. Vielfach erscheinen sie dort als Hintergrundbestandteil gängiger Textfunktionen (z.B. find(), isin(), replace()) oder als Ersatzzeichen (Wildcards). In „Rohform“ angewendet, können reguläre Ausdrücke eine hohen Grad an Komplexität aufweisen. Obwohl es sich um eine sehr effiziente Unterstützungstechnik handelt, wird sie in der Praxis eher selten eingesetzt.

Die Elemente der Beschreibungssprache „Reguläre Ausdrücke“ beinhaltet Festlegungen zu einsetzbaren Zeichenklassen, logischen Verknüpfungstechniken und Quantifizierungen. Der Norm entsprechend bestehen reguläre Ausdrücke aus alphabetischen Zeichen (z.B. Ziffern, Buchstaben, Sonderzeichen, Steuerzeichen), die in Kombination mit sogenannten Metazeichen „[] () { } | ? + - * ^ \$ \“ beliebige Muster beschreiben. Hierbei wird die Meta-Eigenschaft eines Zeichens durch einen vorangestellten Backslash (\) aufgehoben. Ansonsten stehen einzelne Zeichen des Alphabets stehen für sich selbst.²

Für die folgenden Erläuterungen verwenden wir die an Standards orientierten Funktionen des ImportWizard, die in einer zugehörigen Hilfe ausführlich erläutert werden. Eine Verwendung in anderen Programmen (z.B. ActiveData, ACL, IDEA, Excel) kann variieren. Hier muss die zugehörige Dokumentation bei einem Einsatz beigezogen werden.

- **Anwendung in ImportWizard**

Innerhalb des ImportWizard können reguläre Ausdrücke in Markern und Feldformeln verwendet werden. Reguläre Ausdrücke werden über einen Slash (Schrägstrich) eingeleitet. Es folgen ggf.-Gestaltungsoptionen, ein weiterer Slash (Schrägstrich) und der „reguläre Ausdruck“. Die beiden

² Vergl. Wikipedia, https://de.wikipedia.org/wiki/Regulärer_Ausdruck

Schrägstriche sind obligatorisch, auch wenn keine Optionen angegeben sind. Der erste Schrägstrich muss an der ersten Position stehen, andernfalls wird die Zeichenfolge nicht als regulärer Ausdruck interpretiert.

• **Anwendungsbeispiele (ImportWizard)**

Bei den nachfolgenden Beispielen handelt es sich um Kombinationen aus Zeichenklassen, logischen Verknüpfungen, Verbindungen, Quantifizierungen und Metazeichen. Im Zusammenhang mit ihrem Einsatz im Rahmen fortgeschrittener Importtechnik empfehlen wir dringend, mit Hilfe der Dokumentation eine Zerlegung in Einzelschritte und ein logisches Verständnis der Zusammenhänge zu entwickeln:

Muster	Inhalt	Wirkung	Bemerkung
//[0-9]+[.]+[0-9]+[.]+[0-9]+	Beliebiger Ausdruck aus Ziffern und anschließenden Punkten	31.12.2001	z.B. Datum
//\d{5}\$	Fünf numerische Ziffern in Folge in einer Zeile	12345	z.B. Postleitzahl
//(0[1-9] [12][0-9] 3[01])(\.\ V-)(0[1-9] 1[0-2])(\.\ V-)((?:19 20)\d{2})	Datum mit unterschiedlichen Trennzeichen	12.12.2022	Datum in der Form (TT MM JJJJ) mit unterschiedlichen Trennzeichen: ". / -"
//-?(\d{1,3}(\. \,))*(\d{1,3},\d{2})	Beliebig formatierte Beträge mit kommagetrennter Dezimalstelle	12.364,45	
//(?:<=Rechnung).+	Beliebiger Text in einer Zeile <u>rechts</u> neben dem Suchbegriff		Textbegrenzung
//.(?=\sRechnung)	Beliebiger Text in einer Zeile <u>links</u> neben dem Suchbegriff		Textbegrenzung
//(?:<=Rechnung\s).+(?=\sAusgleich)	Beliebiger Text in einer Zeile zwischen zwei Suchbegriffen		Textbegrenzung
//([A-Z]+\s{1,4})+([A-Z]{2})+(\d{10,14})	Vierstelliger Text mit Leerzeichen und zweistelliger Text mit 10 bis 14 Ziffern	IBAN DE39127887790	Selektion
//[A-Za-z]{5}[:]\s?[A-Za-z]{2}?[0-9]{9,14}	Fünfstelliger Text mit anschließendem Doppelpunkt, Leerzeichen, zwei Großbuchstaben und neun bis vierzehn Zahlen	UsTID: DE1234555556	Selektion und Textbegrenzung

Abb. 38 Anwendung regulärer Ausdrücke bei komplexen Suchmustern innerhalb des ImportWizard

Die aufgeführten Beispiele lassen sich beliebig anpassen und erweitern. Bei deren Gestaltung helfen die zahlreichen im Internet auffindbaren „RegEx-Tester“ (u.a. <https://www.regexpal.com> oder <https://pythex.org>), welche auch mit komplexen Textbeispielen gefüllt werden können.

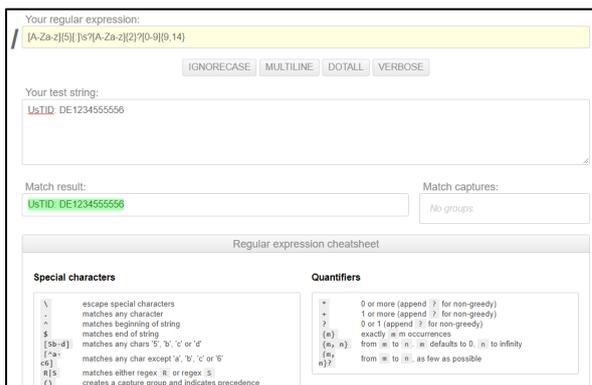


Abb. 39 Internetbasierter RegEx-Tester (pythex) zur Vorab-Kontrolle komplexer Ausdrücke

4 Unsere Prüfungen und Seminare

Die Digitalisierung von Prüfungsaktivitäten führt zu einer erheblichen Spreizung bei Revisionsbereichen, die sich mit entsprechenden Techniken erstmals auseinandersetzen und Innovatoren, welche Unterstützung bei umfassenden Automatisierungen sowie KI-Einsatz erbitten. Erstere möchten einen möglichen Rückstand besonders schnell aufholen und letztere gilt es oft vor überzogenen Erwartungen zu schützen. Es ist nicht immer einfach, diese Fäden - neben der bei uns immer *im Vordergrund stehen Prüfungsarbeit* - zusammenzubinden.

Wir möchten Sie mit unserer Expertise gerne begleiten, bitten in diesem Zusammenhang jedoch um Verständnis, wenn wir bei unseren überwiegend individuell angepassten Seminaren wirklich *seriöse Mitwirkungsanstrengungen* einfordern und zukünftig Vorkenntnisse stärker berücksichtigen. In unserem neuen Seminarkalender haben wir daher das Ihnen bekannte Seminarprogramm umfassend überarbeitet und insbesondere auf unmittelbar praktische Erfordernisse sowie zukunftsgerichtete Aspekte ausgerichtet. Wir werden Sie in Kürze hierzu informieren.

Wir möchten an dieser Stelle auch noch einmal darauf verweisen, dass wir (unabhängig von den verwendeten Programmen) Anregungen für *allgemein interessante prüferische Problemstellungen* gerne aufgreifen und versuchen, praktische digitale Lösungen für den kollegialen Austausch über diese Plattform zu entwerfen. Wir freuen uns in diesem Zusammenhang über jede Kontaktaufnahme und Problembeschreibung.

Zum Abschluss...

bedanken wir uns wieder für Ihr Interesse, Ihre Fragen, Anregungen und die zahlreichen neuen Anmeldungen zu diesem Informationsdienst. Zusätzliche Informationen zu den dargestellten Prüfhilfen finden Sie auf unseren Internet-Seiten:

www.odenthal-auditsoftware.de

www.roger-odenthal.de

Für Anregungen, Rückfragen und Hinweise erreichen Sie uns gerne unter den angegebenen Kontaktdaten.

Mit freundlichen Grüßen


Roger Odenthal


Ute Seeber

