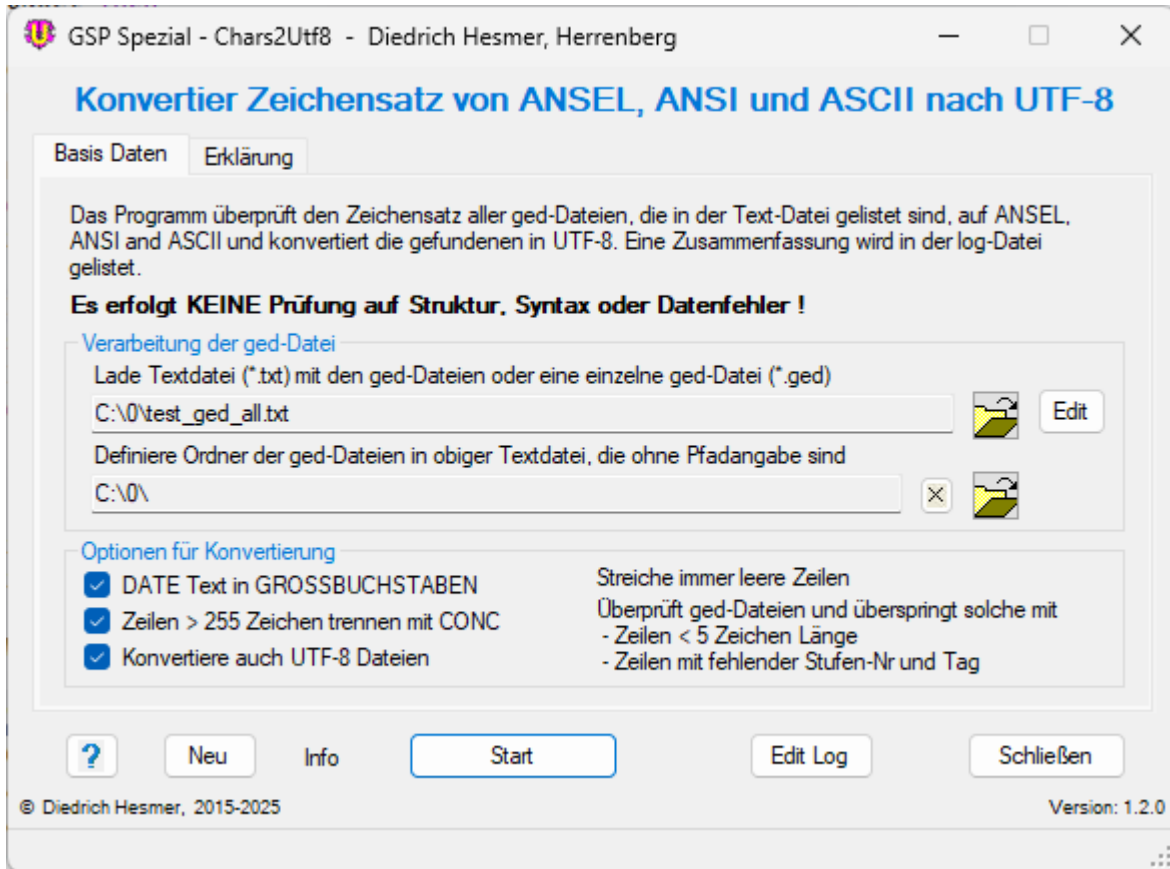




Konvertierung in UTF-8 Zeichensatz



Funktionalität

UTF-8 ist der bevorzugte Zeichensatz für ged-Dateien. Dieser Zeichensatz erlaubt eine unproblematische Darstellung internationaler Zeichen. Einige Programme unterstützen diesen Zeichensatz nicht bei ihrem Export, andere erfordern UTF-8 kodierte Dateien im Import. Neben UTF-8 erlaubt der GEDCOM Standard Entwurf 5.5.1 die Zeichensätze ASCII, ANSEL und UNICODE. Der ANSI Zeichensatz ist nicht erlaubt, allerdings noch weit verbreitet.

Dieses Programm überprüft die aktuelle Kodierung der ged-Datei und konvertiert Dateien mit ANSEL, ANSI oder ASCII in UTF-8 Zeichensatz. In einem Programmablauf kann eine einzelne ged-Datei oder eine ganze Gruppe von ged-Dateien verarbeitet werden.

Zusätzlich zur Konvertierung erfolgt folgendes:

- Leerzeilen werden entfernt.
- Überprüfung ob der HEADER Datensatz vorhanden ist und die aktuell eingetragene CHARS Kodierung.
- Überprüfung auf zu kurze Zeilen und das Vorhandensein von Stufen-Nr und Tag in jeder Zeile.
- Optional können Zeilen > 255 Zeichen aufgetrennt werden und der Text nach dem DATE Tag in GROSSBUCHSTABEN umgewandelt werden.

Durch **Konvertiere auch UTF-8 Dateien** können optional auch ged-Dateien, die bereits UTF-8 kodiert sind, modifiziert werden. Zusätzlich zur ursprünglichen ged-Datei wird eine *_u.ged-Datei erzeugt. Dies ermöglicht eine kontinuierliche Nutzung des Programms "BatchRun".

Die konvertierte Datei wird unter dem Dateinamen der Original Datei mit einem angehängten "_u" gespeichert (xxx.ged wird zu xxx_u.ged). Die Original Datei verbleibt unverändert. Die neue Datei wird im Ordner der Original



Datei gespeichert. Zusätzlich wird eine log-Datei erstellt, die mit jedem Text-Editor geöffnet werden kann. Sie enthält:

- Statistiken über die Konvertierung.
- Gefundene Fehler mit Angabe des Dateinamens und Zeilen-Nummer, sofern möglich.

Beschreibung

Nach Start des Programms erscheint obiger Hauptschirm.

- Es ist zwingend erforderlich, die zu konvertierende ged-Datei(en) zu definieren. Dazu gibt es 2 Möglichkeiten:

- In einer Text-Datei können beliebig viele Dateien definiert werden, je Zeile 1 Datei in dem Format:
 - Voller_Pfad\Dateiname.ged, wobei Voller_Pfad mit dem Laufwerksbuchstaben beginnen muss.
 - Dateiname.ged - hier muss der Volle_Pfad zu dieser Datei im nächsten Feld definiert werden.

Diese Datei muss ".txt" als Dateierweiterung haben. Siehe im Bild unter "Load text file ...". Alle gelistete ged-Dateien werden konvertiert.

Beispiel einer Textdatei mit 4 ged-Dateien:

```
test1.GED
1\die_af\test2.GED
C:\OFB\1\paf\cook\Wilk.ged
C:\OFB\1\die_af\sour\af_dh_td.ged
Zeile 1 + 2 haben keinen Voller_Pfad.
```

- Direkte Angabe einer einzelnen ged-Datei.
Diese Datei muss ".ged" als Dateierweiterung haben und muss ebenfalls geladen werden über "Load text file ...". Nur diese ged-Datei wird konvertiert.
- Optional kann ein Pfad (Ordner) festgelegt werden, um eine Voller_Pfad für Dateien ohne Voller_Pfad zu erstellen.
Nur 1 solcher Pfad kann definiert werden. Diese Pfadangabe wird vor den Dateiangaben gesetzt, die kein Doppelpunkt (:) an der 2. Position haben (Zeile 1 + 2 haben dort keinen Doppelpunkt). Ist der Pfad "C:\OFB\3\1\" im obigen Schirm definiert, so werden die 4 Zeilen aus dem Beispiel geändert in:

```
C:\OFB\3\1\test1.GED
C:\OFB\3\1\1\die_af\test2.GED
C:\OFB\1\paf\cook\Wilk.ged
C:\OFB\1\die_af\sour\af_dh_td.ged
```
- Jede ged-Datei wird während der Konvertierung auf oben beschriebene Fehler geprüft. 2 weitere Optionen zur Korrektur von ged-Dateien sind verfügbar:
 - Der Text nach dem DATE Tag kann in GROSSBUCHSTABEN umgewandelt werden.
 - Zeilen mit einer Länge > 255 Zeichen können unter Verwendung des CONCatenate Tag aufgeteilt werden.

Folgende Schaltknöpfe sind vorhanden

- [Symbol mit öffnenden Ordner] Lade Datei bzw. Ordner/Pfad.
- [Edit] Editiert die ged-Datei mit dem Standard Text Editor.
- [X] Löscht den Inhalt der Text box.
- [?] Aufruf Programmbeschreibung.
- [Neu] Löscht alle Definitionen.
- [Info] Zeigt den Ordner, in der die ini-Datei gespeichert ist (enthält alle Einstellungen während des letzten "Schließen").



- [Start] Startet die Konvertierung.
- [Edit Log] Öffnet die log-Datei durch den Text Editor.
- [Schließen] Schreibt die ini-Datei mit allen Einstellungen und beendet das Programm. Beendigung über das kleine "x" oben rechts am Bildschirm schreibt keine ini-Datei.)

Auszug aus einer Log-Datei:

```
Erstellt am 11.02.2015 21:38:49 - Chars2Utf8 version: 1.0.2
Basis: "c:\1\test_ged_all.txt" vom 10.02.2015 19:43:27
Folgende Dateien nicht gefunden, keine ged-Dateien oder enden mit "_u.ged":
C:\Ofb\3\1\bk\test_f\test2_u.ged
C:\Ofb\3\1\die_af\af-hes-paf_u.ged
C:\Ofb\3\1\die_af\af_dh_test-vn_u.ged
C:\Ofb\3\1\die_af\sour\af_dh_neu_sor_u.ged
...
C:\Ofb\3\1\paf\sour\willy50_u.ged
574 ged-Dateien gefunden in Textdatei
11.02.2015 21:38:49: Start Verarbeitung
Folgende Dateien wegen Fehler nicht konvertiert:
Zeile Nr. 9390 "1 " zu kurz -> Datei "C:\Ofb\3\1\ages\richt\hd-ofb9-2014.ged"
Zeile Nr. 261 "Bad Pyrmon..." Ziffer + Tag fehlen -> Datei "C:\Ofb\3\1\bk\spiel\2014 (2).ged"
Zeile Nr. 315 "1 -DET VOG..." Ziffer + Tag fehlen -> Datei "C:\Ofb\3\1\easy_tree\wink\1-4test-
vf.ged"
Zeile Nr. 22 "3_ZUS" Ziffer + Tag fehlen -> Datei "C:\Ofb\3\1\gen_p\marzi\miesen.ged"
Unbekannte Kodierung "" im HEADER (Kopf) -> Datei "C:\Ofb\3\1\gramps\paktig\hesmer_test.ged"
Zeile Nr. 58 "2 07 MAR 1..." Ziffer + Tag fehlen -> Datei "C:\Ofb\3\1\paf\karl\ble83.ged"
Zeile Nr. 366 "... " zu kurz -> Datei "C:\Ofb\3\1\turck\kb_meinerz.ged"
"0 HEAD" Zeile fehlt -> Datei "C:\Ofb\3\1\webtree\utischill\uti_familytree_test-reduzierer_red.ged"

11.02.2015 21:45:37: Konvertierung abgeschlossen
253 Dateien übersprungen - bereits UFT-8
271 ANSI Dateien konvertiert
46 ANSEL Dateien konvertiert
2 ASCII Dateien konvertiert
46.341.153 Zeilen gelesen - total
46.397.016 Zeilen geschrieben - total
```

Der obere Teil listet die Programme, die nicht gefunden wurden, die keine ged-Dateien sind oder bereits mit _u.ged enden. Diese werden NICHT konvertiert.

Der mittlere Teil listet fehlerhafte Dateien bzgl. fehlendem HEADER, unbekannter oder nicht unterstützter CHARS Kodierung, zu kurzer Zeilenlänge oder fehlender Stufen-Nr. und Tag. Diese werden übersprungen. Der untere Teil listet eine Zusammenfassung der Konvertierung.

Beachte:

Während der Verarbeitung der Daten erfolgt **keine** Überprüfung korrekter GEDCOM Strukturen und Syntax, Gültigkeit und Konsistenz der Daten der ged-Datei. Dies kann vorher mit Hilfe anderer GSP Programme sichergestellt werden.

Versionen

(N = Neu, K = korrigiert, M = modifiziert)

- 1.0.0 - 20.02.2015 - N: Erstveröffentlichung
- 1.0.1 - 25.02.2015 - N: Deutsche Oberfläche
- 1.1.0 - 13.03.2015 - M: Dateiname _u
- 1.1.1 - 25.02.2017 - M: Datei öffnen standardisiert
- 1.1.2 - 01.07.2017 - M: Lizenzfunktion
- 1.1.3 - 10.05.2018 - N: "?" Aufruf Beschreibung
- 1.2.0 - 22.03.2025 - N: Option "Konvertiere auch UTF-8 Dateien"