



GEKo

der Gedcom-Encoding-Konverter

GEKo dient der Konvertierung von Gedcomdateien.

Viele genealogische Programme nutzen Gedcomdateien zum Austausch von genealogischen Familiendaten. Die zur Codierung von Umlauten und Sonderzeichen vorgesehenen Encodings sind in der aktuellen Gedcomdefinition 5.5 vorgegeben. Die geplante Version Gedcom 5.5.1 liegt in der Fassung vom 2. Oktober 1999 vor, wurde aber bisher (Januar 2006) nicht verabschiedet. Es scheint es keine Bestrebungen zu geben, an diesem Zustand etwas zu ändern. Momentan bedeutet das, es stehen folgende zulässige Codierungen zur Verfügung:

ASCII

... die schlechteste Wahl. Diese Codierung erlaubt keine Umlaute und Sonderzeichen, da dieser Zeichenvorrat mit 7-bit pro Zeichen auskommen muss. Es stehen somit nur 127 Zeichen zur Wahl. Zeichen mit einem Wert größer als 127 wandelt GEKo in ein Fragezeichen um.

ANSEL

... kann viele Sonderzeichen codieren, auch solche, die in keiner Sprache genutzt werden. Leider ist diese Codierung außerhalb von Gedcom recht unüblich. Viele Genealogieprogramme scheinen aber recht gut damit zurecht zu kommen. Es empfiehlt sich zu prüfen, ob wirklich alle Sonderzeichen vom importierenden Programm richtig übernommen werden. Codierbar sind ausnahmslos Zeichen, die auf einer Grundform lateinischer Zeichen basieren. Somit sind weder kyrillische, hebräische oder diverse asiatische Zeichen darstellbar.

Leider kennt GEKo nicht alle mit ANSEL codierbaren Zeichen. Sollten Sie Zeichen finden, die nicht (oder nicht richtig) codiert werden, bitte ich um einen entsprechenden Hinweis.

UNICODE

... erlaubt (neben UTF-8) die meisten Zeichen. Leider hat sich UNICODE noch nicht so verbreitet, wie es wünschenswert wäre. Mit UNICODE sind innerhalb der Gedcomdatei selbst unterschiedlichste Zeichen wie chinesisch, hebräisch und kyrillisch möglich. Leider unterstützen nicht alle Programme diese Codierung.

UTF-8

... eine Abwandlung von Unicode. UTF-8 ist erst in Gedcom 5.5.1 Draft¹ definiert und gehört somit genaugenommen nicht zum Standard von Gedcom 5.5. Da UTF-8 jedoch platzsparender als Unicode ist und zudem noch in den ersten 255 Zeichen weitestgehend dem Windows Zeichenvorrat entspricht, setzt es sich immer mehr durch. Meiner Einschätzung nach unterstützen mehr Programme UTF-8 als UNICODE. Es gibt sogar Programme, die UTF-8 importieren, UNICODE jedoch nicht. Umgekehrt scheinen alle Programme, die UNICODE importieren können auch UTF-8 zu importieren.

¹ „Draft“ ist die englische Bezeichnung für einen Entwurf. Dieser Entwurf ist nicht als abschließend fertiger Standard freigegeben worden.

NICHT ERLAUBT

... sind alle anderen Codierungen. Genaugenommen dürften sich Gedcomdateien nicht als solche bezeichnen, wenn eine andere Codierung angegeben ist. So finden sich in vielen Gedcom-Dateien Angaben wie ANSI, DOS, WINDOWS, PC, IBM, etc. Gerade diese sind aber in der Gedcomdefinition verboten (Gedcom 5.5. Definition):

„The IBMPC character set is not allowed. This character set cannot be interpreted properly without knowing which code page the sender was using.“

Zu deutsch etwa:

„Der IBMPC Zeichenvorrat ist nicht erlaubt. Dieser Zeichenvorrat kann nicht sicher interpretiert werden, wenn an nicht weiß, welche Code-Page¹ der Sender benutzt hat.“

Genau aus diesem Grund gibt es GEKo. Mit diesem Programm können Sie viele Codierungen die im Umlauf sind in eine dem Gedcomstandard entsprechende Codierung wandeln. Der Inhalt und der Aufbau der Gedcomdatei wird dadurch nicht verändert. Es werden lediglich die Umlaute und Sonderzeichen (alle Zeichen mit einem Bytewert über 127) sowie die Zeile mit der Kodierungsangabe (z. B. 0 CHAR ANSEL) umgewandelt. Sollten Sie in den Einstellungen noch die Zeilenenden oder ein Byteordermark (nur bei Unicode) verändern oder hinzufügen wollen, wird das natürlich ebenfalls geändert.

Installation

GEKo kann direkt ohne Installation gestartet werden. Kopieren Sie die Programmdatei lediglich an den gewünschten Ort. Bei vielen Betriebssystemen gibt es bestimmte Vorgaben, in welchem Verzeichnis sich Programme befinden sollten. Zwingend erforderlich ist dieses Verzeichnis für die korrekte Funktion von GEKo nicht.

Nach dem Beenden von GEKo schreibt das Programm diverse Informationen (Fensterpositon und Einstellungen) in die Datei „geko.inf“ im Preferences-Ordner. Dort wird ggf. die Datei „GEKo-UserCodes.txt“ gesichert. Sie enthält Zeichenumsetzungen zur Wandlung von/zu Ansel, die Sie selber angegeben haben.

Der Preferences-Ordner befindet sich unter

Linux: das Homeverzeichnis des aktuellen Benutzers

Mac OS X: der Ordner „Preferences“ im Systemordner

Windows: C:\Dokumente und Einstellungen\[Anwendername]\Lokale Einstellungen\Anwendungsdaten\

¹ Der Zeichenvorrat von Ansi (IBMPC) unterscheidet verschiedene länderspezifische Gruppen von Sonderzeichen. Diese bezeichnet man als Code-Page. Zeichen aus unterschiedlichen Code-Pages lassen sich nicht innerhalb einer Datei mischen.

Deinstallation

Um GEKO wieder zu entfernen, löschen Sie einfach die Programmdatei. Sofern Sie GEKO bereits benutzt haben, befinden sich noch die Datei „geko.inf“ und ggf. „GEKO-UserCodes.txt“ im Preferences-Ordner (s.o.), auch diese können Sie einfach löschen. Sie können Sie aber dort auch lassen, sie haben keinen Einfluss auf Ihr System.

WICHTIG!

Die Risiken der Nutzung von GEKO unterliegen einzig und allein dem Nutzer. Ich übernehme keinerlei Haftung aus Schäden die durch die Nutzung oder ein Fehlverhalten von GEKO entstehen.

Wie benutze ich GEKO?

Starten Sie zuerst das Programm GEKO durch einen Doppelklick auf die Programmdatei. Nachdem Sie den Infodialog geschlossen haben, können Sie eine Gedcomdatei in das Programmfenster ziehen oder über den Menüpunkt „Gedcomdatei öffnen“ eben diese öffnen. GEKO analysiert die Gedcomdatei und versucht die Zeichencodierung zu bestimmen. Welche Kodierung erkannt wurde, wird im Auswahlmenü unten links (hinter „Originaldatei:“) angezeigt.

Im unteren Teil des Fensters wird die Gedcomdatei angezeigt. Hier können Sie sehen, wie die Originaldatei aussieht. Darüber befindet sich ein Bereich, der mehrere Unterpunkte bietet. Sie werden als erstes die Headerdaten angezeigt bekommen. Darin können diverse Informationen enthalten sein., So z. B. der Name und die Versionsnummer des exportierenden Programms, die Adresse dessen, der die Gedcomdatei erstellt hat, etc.

Des weiteren können sie sich im Unterpunkt „Statistik“ ein Übersicht über die Anzahl der enthaltenen Personen, Familien, Bemerkungsrecords, Quellenangaben, etc. ansehen.

Abschließend lohnt sich ein Blick auf die Bemerkung. Wenn diese Umlaute enthält, können Sie daran schon erkennen, ob GEKO die original Textcodierung richtig erkannt hat. Sollte das mal nicht der Fall sein, Stellen Sie die richtige Codierung im Auswahlmenü „Originaldatei:“ passend ein.

Vor der eigentlichen Konvertierung rauchen Sie jetzt nur vorgeben, welche Codierung Sie für die Ausgabedatei wünschen. Dazu dient das Auswahlmenü unten rechts („Ausgabedatei“). Wenn Sie auch diesen Schritt erledigt haben, brauchen Sie nur noch auf den Button „Konvertieren...“ Klicken. Danach fragt GEKO, wo und unter welchem Namen die konvertierte Datei gesichert werden soll.

Sollten Sie Fragen oder Anregungen zum Programm haben, wenden Sie sich bitte an:

s.metti@gmx.de

Die aktuelle Version finden Sie auf meiner Homepage:

<http://www.familienbande-genealogie.de>

Viel Erfolg,

Stefan Mettenbrink