



Suchen von Gedcom Tag Kombinationen

Util - GedTagSearch - Diedrich Hesmer, Herrenberg

Suchen von Gedcom Tag Kombinationen

Basis Daten | Liste der Dateien | Ergebnisliste | Beispiele

Programm durchsucht ged-Dateien auf definierte Tagkombinationen und Tag-Text-Muster und gibt Anzahl bzw. Zeilen-Nm aus. In wesentlichen Teilen sind Jokerzeichen und Textmuster Vergleiche erlaubt. Einige Beispiele befindet sich im Reiter "Beispiele".

Startverzeichnis, Unterverzeichnisse + Dateiname festlegen

Startverzeichnis: D:\Ofb\1\genea\
Unterverzeichnisse:
Dateinamen: * schließe alle Unter-Ordner ein

Auswahl von Startverzeichnis, Eingabe der Unterverzeichnisse und Dateinamen (ohne Endung) der ged-Dateien. Der Dateiname kann Jokerzeichen enthalten. Mehrere Dateinamen durch ";" trennen.

Tags definieren

| Datensatz | TAG_1 -> 1 Tag [Text] | TAG_2 -> 2 Tag [Text] | TAG_3 -> 3 Tag [Text] | TAG_4 -> 4 Tag Text |
|-----------|-----------------------|-----------------------|-----------------------|---------------------|
| I | 1 DEAT + | | | |
| I | 1 BIRT | 2 DATE - | | |
| I | F 1 *;MARR;CHR | 2 PLAC | 3 _GOV - | |
| I | F 1 * | 2 PLAC * | 3 MAP | 4 LONG W-* |
| | L 1 _GOV - | | | |
| | | | | |

Je Zeile Eintrag von: Suche in INDI/FAM/SOUR/_LOC Datensatz, Stufe-Nr, Tag, ggf. Text für Stufe 1 und ggf. Stufe 2..4. "TAG_1" darf Jokerzeichen * enthalten. "Text" darf die Jokerzeichen * # ? enthalten. Ein einzelnes Zeichen am Ende von "Text" sucht: "%" = fehlender Text, "-" fehlend Tag, "+" mehrl Tag. Ist TAG_2 ein TYPE Tag, so muss TAG_3 ein Stufe-2 Tag sein.

Optionen

Ausgabe Anzahl Ausgabe HEAD.CHAR Ausgabe gefundene Tags GSP ausschließen
 Ausgabe Zeilen-Nm Ausgabe HEAD.SOUR Ausgabe *1** Tags Specials ausschließen
 Nur Zusammenfassung Ausg. Datensätze/Datei Ausgabe * DATE Andere ausschließen
Abbruch bei Funde > 0

Unterdrücke Text
Tag_ASSO -> ASSO

GedTagSearch Steuerdatei
D:\Ofb\1_files\ugts\test-ug.ugts

Neu Info Start Edit Log Schließen

© Diedrich Hesmer, 2015-2022 Version: 2.3.2
Datei "C:\DatVS\VS 2015\Projects\iniFiles\ExclFileGts.txt" eingelesen

Funktionalität

Dieses Programm durchsucht ged-Dateien auf definierte Tag-Text-Muster. Gesucht wird in INDI, FAM, SOUR und/oder _LOC Datensätzen. Ausgegeben werden je nach Option Detailedaten wie die gefundenen Namen/Familien mit oder ohne die gefundenen Tags mit ihren Texten, die Zeilen-Nummern und die Anzahl der Funde. Die Suche kann über bis zu 4 Gedcom Tag Stufen erfolgen.

Außerdem kann gesucht werden nach

- Tags ohne Text
- Fehlende Tags
- Mehrfach vorkommende Tags

Während des Ablaufs wird eine log-Datei erstellt, die mit jedem Text-Editor geöffnet werden kann.

Beschreibung

Nachfolgend sind der Ablauf, die Suchkriterien, die Optionen, die Wirkung der Schaltknöpfe, Beispiele und die log-Datei beschrieben.



Ablauf

Nach Start des Programms erscheint obiger Hauptschirm.

Er besteht aus 3 Reiter zur Festlegung der Selektion, zur Ausgabe der gefundenen Dateien und deren Selektion zur Analyse und zur Ausgabe der Ergebnisse.

In einem Lauf können bis zu 6 Tag Kombinationen in beliebig vielen Dateien gesucht werden. Hierzu ist ein "Startverzeichnis", mindestens 1 Datei (auch "*") und mindestens 1 Tagkombination zu definieren. Neben dem Startordner können auch definierte Unterordner und von denen alle Unterordner durchsucht werden.

- Beim ersten Aufruf sind alle Felder des Startfensters leer bzw. auf ihren Systemwert gesetzt.
- Definition von Startverzeichnis und Dateien.
- Es sind die Suchkriterien für die zu untersuchende ged-Dateien und die Tag-Text-Muster zu definieren.
- Auswahl der Optionen.
- Mit "Start" startet die Suche. Aus dem Startverzeichnis, den Unterverzeichnissen und den ged-Datei Definitionen werden die betreffenden ged-Dateien gesucht und in der "Liste der Dateien" zur Überprüfung und Auswahl aufgelistet.
- Mit "Suchen" beginnt die Überprüfung der ausgewählten Dateien entsprechend den Tag-Muster Definitionen.
- Die Funde werden für jede Datei in der "Ergebnisliste" und in der log-Datei entsprechend den Ausgabe Optionen ausgegeben.
 - Zeilen-Nr, Datensatztyp, Tagkombination, Datensatz, ggf. Name, ggf. Zeilen von weiteren Funden
 - Anzahl gefundener Kombinationen
 - Gefundene Tags mit Anzahl für "1 *" Tags
 - Gesamtzahl gefundener Kombinationen
 - Liste der Dateien ohne Funde
- Die Datei mit dem Dateinamen "GedTagSearch.log" ist im Startordner gespeichert.
- Bei einem erneuten Aufruf sucht das Programm nach einer ini-Datei. Falls gefunden, werden die Felder entsprechend den Einstellungen vom letzten "Schließen" gefüllt. Diese können vor erneutem "Start" beliebig geändert werden.

GedTagSearch Steuerdatei

Die "ugts" Steuerdatei enthält sämtliche im Hauptschirm vorgenommene Einstellungen für die spätere erneute Verwendung.

- Mittels dem Schaltknopf [Symbol mit der Diskette] wird der Speicherort (Verzeichnis) und der Name der Steuerdatei bestimmt und diese als ugts-Datei dort abgelegt.
- Mittels dem Schaltknopf [Symbol mit öffnenden Ordner] kann eine gespeicherte ugts-Datei erneut geladen und die Einstellungen übernommen werden.

Startverzeichnis und Dateien

Ausgehend von einem festzulegenden Startverzeichnis können bestimmte Unterverzeichnisse und ged-Dateien, die zu durchsuchen sind, definiert werden.

- Für die Verzeichnisse kann gewählt werden, ob
 - nur das Startverzeichnis, optional mit allen seinen Unterverzeichnissen untersucht werden soll oder
 - das Startverzeichnis mit zusätzlich definierten Unterverzeichnissen, optional mit deren Unterverzeichnissen untersucht werden soll.
 - Bei der Wahl von mehreren definierten Unterverzeichnissen sind diese durch ";" zu trennen.
Beispiel: Mit "valid;conv" werden vom Startverzeichnis die Unterverzeichnisse "valid" und "conv" untersucht.
- Für die Wahl von Dateien gilt:
 - Es muss hier ein Eintrag erfolgen.



GSP Dienste – GedTagSearch

- Die Dateinamen sind ohne die Erweiterung ".ged" einzutragen.
- Mehrere Dateinamen sind durch ";" zu trennen.
- Es können im Dateinamen die Jokerzeichen * # ? verwendet werden.
Beispiel: "*_u;*_sor" verarbeitet die ged-Dateien, deren Dateinamen mit "_u.ged" und "_sor.ged" enden.
- Wird nur das Jokerzeichen "*" eingegeben, werden alle ged-Dateien analysiert.
- Es können durch GSP-, Special- oder Andere-Programme erstellte ged-Dateien abgewählt werden. Siehe nächster Absatz.

Steuerdatei "ExclFileGts.txt"

Dies ist eine optionale Datei. Sie dient dazu, ged-Dateien, deren Dateiname bestimmte "signifikanten Zeichen" enthalten, aus der Analyse auszuschließen. Diese sind gekennzeichnet durch die letzten 3-5 Zeichen (sog. signifikante Zeichen) im Dateinamen der ged-Datei, also die Zeichen vor ".ged". Beispiel: Die ged-Datei hat den Namen xxx.ged. Der GSP Reduzierer macht daraus ein xxx_red.ged, der GSP Sortierer ein xxx_sor.ged. Damit sind die signifikanten Zeichen "_red" bzw. "_sor". Der Inhalt dieser Steuerdatei überschreibt die Systemwerte. Die Datei ist eine einzeilige Textdatei, die im Verzeichnis der ini-Dateien gespeichert werden muss. Sie enthält die letzten signifikanten Zeichen aus den Namen der auszuschließenden ged-Dateien.

Aufbau der Zeile:

- Sie enthält 3 Gruppen, getrennt durch Semikolon ";": **GSP-Gruppe;Specials-Gruppe;Andere-Gruppe**. Damit sind 2 Semikolon erforderlich. Wird eine Gruppe nicht definiert, werden für diese die Systemwerte genommen.
- Jede Gruppe enthält die signifikanten Zeichen der auszuschließenden ged-Dateien, jeweils durch Komma "," getrennt.
- Beispieldatei (in einer Zeile):
`_a,_mer,_num,_red,_sm,_sor,_ss,_syn,_tmp,_u;_adt,_jf,_dif,_ehe,_f5o,_g2h,_n2g,_n2p,_n2pz,_pe2a,_r2g,_shw,_2t1;_dir,_c2g,_mgf,_mnt,_rep`

Selbstverständlich kann auch jede beliebige andere Text, der das Ende eines Namens der ged-Datei entsprechen, als signifikante Zeichen festgelegt werden.

Suchkriterien

Die Suchkriterien bestimmen die Suche nach den betreffenden Tag-Text-Kombinationen. Fehlerhafte Eingaben sind pink hinterlegt und müssen korrigiert werden.

Bis zu 4 unterschiedliche Kombinationen von Tags mit/ohne Texte können definiert werden. Dabei gilt:

- Auswahl der zu durchsuchenden Datensätze: INDI, FAM, SOUR, _LOC. Bis zu 2 dieser Datensätze können für eine Definitionszeile gewählt werden.
- Festlegung des Stufe-1 Tags (TAG_1) in der Form "1 Tagname" bzw., sofern das Tag auch Text enthält, "1 Tagname Text". Als Sonderfall ist "1 *" erlaubt. Hierbei werden alle Stufe-1 Tags durchsucht. Optional können dafür die gefundenen Tags und deren Anzahl ausgegeben werden.
Besonderheit: Beim Sonderfall "1 *" können Stufe 1 Tags, die NICHT analysiert werden sollen, abgewählt werden. Hierzu sind die Tagnamen nach dem "1 *" durch Semikolon ";" und Tagname anzugeben. Mehrere Tagnamen sind durch Semikolon ";" zu trennen. Beispiel: "1 *;BIRT;DEAT;MARR".
- Festlegung des Stufe-2 (TAG_2), Stufe-3 Tags (TAG_3) und Stufe-4 Tags (TAG_4) in der Form "x Tagname" bzw., sofern das Tag auch Text enthält, "x Tagname Text". Diese Felder sind optional und normalerweise mit Text zu definieren.
Beachte: Ist die Stufe-2 ein TYPE Tag, so muss das Feld von TAG_3 ein Stufe-2 Tag sein oder leer bleiben, TAG_4 muss ebenfalls leer bleiben.
- Der "Tagname" muss den kompletten Tagnamen enthalten. Alle "Texte" der Stufe 1 bis 4 dürfen die Jokerzeichen * ? # enthalten. Außerdem sind Textmuster Vergleiche erlaubt, wie im GSP Handbuch Teil 1, im Anhang unter "Zeichenfolgevergleiche" beschrieben. Texte müssen bzgl. GROSS/kleinschrift dem Original entsprechen.



GSP Dienste – GedTagSearch

Beachte: Um nach * # ? zu suchen, müssen diese in eckige Klammern gesetzt werden, z.B. "[?]", diese Felder können leer bleiben, wenn die HEAD Optionen gewählt sind.

Besondere Steuerzeichen:

- **Fehlender Text:** Wird bei den Stufen 1-4 als "Text" ein einzelnes "%" Zeichen eingegeben, so wird das entsprechende Tag ohne Text gesucht. Beispiel: "1 SEX %".
- **Fehlendes Tag:** Wird bei den Stufen 1-4 als "Text" ein einzelnes "-" Zeichen eingegeben, so werden die Datensätze gesucht, bei denen das entsprechende Tag fehlt.

Beachte: "1 * -" ist nicht erlaubt, in jeder Zeile darf nur 1 "-" oder "+" vorkommen.

| Datensatz | TAG_1 -> 1 Tag [Text] | TAG_2 -> 2 Tag [Text] | TAG_3 -> 3 Tag Text | TAG_4 -> 4 Tag Text |
|-----------|-----------------------|-----------------------|---------------------|---------------------|
| I | 1 DEAT + | | | |
| I | 1 BIRT | 2 DATE - | | |
| I | 1 *:MARR;CHR | 2 PLAC | 3 _GOV - | |
| I | 1 * | 2 PLAC * | 3 MAP | 4 LONG W-* |
| L | 1 _GOV - | | | |

Im obigen Beispiel wird gesucht in Zeile

- 1 in INDI Datensätzen nach mehrfachen DEAT Tags
 - 2 in INDI nach fehlenden DATE Tags unter BIRT
 - 3 in INDI und FAM nach fehlenden _GOV Tags unter PLAC unter allen Stufe 1 Tags außer MARR und CHR
 - 4 in INDI und FAM für alle Stufe-1 Tags bei allen PLAC Tags unter MAP nach LONG Texten gesucht, die mit "W-" beginnen. Das Jokerzeichen "*" ist wichtig, damit das "-" nicht als Steuerzeichen erkannt wird.
 - 5 in _LOC nach fehlenden _GOV Tags.
- **Mehrfaches Tag:** Wird bei den Stufen 1-3 als "Text" ein einzelnes "+" Zeichen eingegeben, so werden die Datensätze gesucht, bei denen das entsprechende Tag mehrfach vorhanden ist. **Beachte:** In jeder Zeile darf nur 1 "-" oder "+" vorkommen.

| Datensatz | TAG_1 -> 1 Tag [Text] | TAG_2 -> 2 Tag [Text] | TAG_3 -> 3 Tag Text |
|-----------|-----------------------|-----------------------|---------------------|
| I | 1 * +:OCCU;OBJE | | |
| F | 1 MARR + | | |
| I | 1 DEAT | 2 PLAC | 3 _LOC - |
| I | 1 EVEN | 2 TYPE * | |

Im obigen Beispiel wird gesucht in Zeile

- 1 in INDI und FAM Datensätzen nach mehrfachen Stufe-1 Tags außer nach OCCU und OBJE
- 2 in FAM nach mehrfachen MARR Tags
- 3 in INDI nach fehlenden _LOC Tags unter DEAT.PLAC
- 4 in INDI nach mehrfachen EVEN Tags mit allen TYPE Texten.

Ein Fund ist dann gegeben, wenn in der ged-Datei eine Übereinstimmung mit den definierten Kriterien gefunden wurde.

Optionen

Als Optionen stehen zur Verfügung:

- Alternative Optionen – eine muss gewählt werden:
 - **Ausgabe Anzahl** der gefundenen Tagkombinationen in den einzelnen Dateien und Gesamt. Wird auch bei gefundenen Zeilen ausgegeben.
 - **Ausgabe Zeilen-Nrn** mit den Tagkombinationen -> Zeilen-Nr, Tagkombination, Datensatz, ggf. Name der Person.
 - **Nur Zusammenfassung** gibt nur die Gesamtsummen aus.
- Ausgabe von Name des Genealogie-Programms (**HEAD.SOUR**) und Zeichensatz (**HEAD.CHAR**) zur jeweiligen ged-Datei.



- **Ausg. Datensätze/Datei**
- **Ausgabe gefundene Tags** gibt die Tagkombinationen mit den Texten aus.
- **Ausgabe "1 *" Tags** Ausgabe der gefundenen Tags und deren Anzahl für jede Datei und Gesamt.
- **Ausgabe * DATE** Ausgabe des Geburts- bzw. Taufdatums zum Namen.
- **Unterdrücke Text** gibt nur die Tagnamen ohne ihren Text aus.
- **Tag_ASSO -> ASSO** veranlasst beim Lesen eine Umwandlung des Tags "_ASSO" in "ASSO".
- **GSP / Specials / Andere ausschließen** schließt diese Dateitypen aus.
- **Abbruch bei Funde** > nnn veranlasst ein Abbruch bei mehr Funde als angegeben. Wird "0" angegeben, wird alles ermittelt.

Schaltknöpfe

Zusätzlich zu den oben beschriebenen Schaltknöpfen für die ugts-Steuerdatei existieren:

- [Symbol mit öffnenden Ordner] Festlegung eines Startverzeichnisses für die Suche nach ged-Dateien.
- [x] rechts neben den Zeilen löscht alle Einträge der Zeile
- [x] unterhalb der Tagdefinitionen löscht alle Einträge der Tagdefinitionen.
- [^] erlaubt ein Verschieben einer der 5 Zeilen (2..6) um 1 Position nach oben. Dazu ist in der zu verschiebenden Zeile das TAG_1 Textfenster zu markieren und danach dieser Schaltknopf zu betätigen. Der Inhalt diese Zeile wird mit der darüberstehenden Zeile ausgetauscht.
- [?] Aufruf Programmbeschreibung.
- [Neu] Löscht alle Definitionen im Fenster.
- [Info] Zeigt den Ordner, in der die ini-Datei gespeichert ist (enthält alle Einstellungen während des letzten "Schließen").
- [Start] Startet die Verarbeitung.
- [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.
- [Suchen] im Reiter "Liste der Dateien" startet den Suchlauf.

Beispiele

Beispiele dazu finden Sie im Reiter " Beispiele".

Auszug aus einer Log-Datei

Sie enthält zusammenfassende Ergebnisse der Analyse und ggf. für jeden Fund:

- Zeilen-Nr. und Suchstruktur.
- Datensatz Id.
- Namen der Person des betreffenden Datensatzes.
- Aktuelle Tags mit Merkmalstext, bei Mehrfachtags die weiteren Zeilen-Nrn.
- Zusammenfassung der Funde einschl. die Dateinamen ohne Fund.

Hier ein Auszug einer Datei:

```
05.11.2017 16:16:02 Start Scan - GedTagSearch Version: 1.2.0
Zeile-Nr: Tagkombination -> Datensatz Id
*** Datei: "C:\Ofb\1\pgm_test\ofb-orig\ofb_dh.GED" -> SOUR: diverse, CHAR: ANSI, INDI: 23, FAM: 10
    56: 1 ADOP.2 AGE -> @I1@ [Dagmar_Erika /Wfrau/] 1 ADOP 2 AGE 3 Yrs
    558: 1 *.2 HUSB.3 AGE * -> @F1@ 1 ENGA 2 HUSB 3 AGE 23 y
    1 x FAM: 1 *.2 HUSB.3 AGE *
    1 x ENGA
    1 x IND: 1 ADOP.2 AGE

Total gefunden in 1 Datei(en):
    1 x FAM: 1 *.2 HUSB.3 AGE *
    1 x ENGA
    1 x IND: 1 ADOP.2 AGE
    0 x IND: 1 SEX U
Gesamtzahl Datensätze - INDI: 23, FAM: 10

Keine Funde in 2 Datei(en):
```



C:\Ofb\1\pgm_test\ofb-orig\AF_dh_td.GED
C:\Ofb\1\pgm_test\ofb-orig\AF_dh_test-VN.ged
05.11.2017 16:16:02 Analyse der Dateien abgeschlossen

Versionen

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

- 1.0.0 – 25.02.2017 – N: Erstveröffentlichung
- 1.0.1 – 01.07.2017 – M: Lizenzfunktion
- 1.0.2 – 01.08.2017 – N: Limit für Abbruch
- 1.1.0 – 10.10.2017 – N: Ausgabe gefundener "*" 1" Tags
- 1.2.0 – 20.11.2017 – N: Checklist für Dateien, Liste der Dateien ohne Fund, Stufe-2 TYPE
- 1.2.1 – 19.02.2018 – M: Prüfung Existenz von Text Stufe-2/3
- 1.2.2 – 08.05.2018 – N: "?" Aufruf Beschreibung
- 1.3.0 – 03.03.2020 – N: Steuerdatei "ExclFileGts"
- 1.4.0 – 31.03.2020 – N: Option "Unterdrücke Text"
- 1.5.0 – 30.10.2020 – N: "%" Zeichen als Text zeigt Tags mit fehlendem Text
- 1.6.0 – 01.12.2020 – N: Angabe zu übergehender Stufe-1 Tags bei "1 *"
- 2.0.0 – 25.12.2020 – N: "-" Zeichen als Text sucht fehlende Tags
- 2.1.0 – 20.01.2021 – N: "+" Zeichen als Text sucht mehrfache Tags, ugts-Steuerdatei
M: "" in "-" getauscht
- 2.2.0 – 01.02.2021 – N: Die Datensätze SOUR und _LOC können ebenfalls durchsucht werden.
- 2.3.0 – 18.12.2021 – N: Anzahl Tag Kombinationen 4 > 6, zusätzlich TAG_4
- 2.3.1 – 30.01.2022 – M: UTF-8 Konvertierung
- 2.3.2 – 04.04.2022 – N: Ausgabeoption für Geburts/Taufdatum