



Search of Gedcom Tag Combinations

Util - GedTagSearch - Diedrich Hesmer, Herrenberg

Search of Ged Tag Combinations

Base Data | List of Files | List of Results | Examples

Program scans ged-files for defined tag combinations and tag-text pattern and displays quantities resp. line numbers. In most cases wildcard chars and text pattern comparisons are allowed. Some examples are at tab "Examples" available.

Start-folder, Sub-folder + File name definition

Start-folder: D:\Ofb\1\genea\

Sub-folders: multiple separated by ";"

File names: * ☐ include all Sub-folders

Selection of start-folder, input of sub-folders and file names (w/o extension) of ged-files.
The file name may include wildcard characters. Multiple files separated by ";".

Tags definitions

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

For each row define: Search of INDI/FAM/SOUR/_LOC record, level-nbr., tag, i.a. Text for level 1 and i.a. level 2..4.
"TAG_1" may include the wildcard character *. "Text" may include the wildcard characters * # ? . A single char at end of "Text" searches for: "% missing text, " missing tag, "+" multiple tags. If TAG_2 is a TYPE tag, then TAG_3 must be a level-2 tag.

Options

<input type="radio"/> Output quantity found	<input checked="" type="checkbox"/> Output HEAD.CHAR	<input checked="" type="checkbox"/> Output tags found	<input type="checkbox"/> GSP-Files excluded
<input checked="" type="radio"/> Output line numbers	<input checked="" type="checkbox"/> Output HEAD.SOUR	<input checked="" type="checkbox"/> Output "1 *** tags	<input checked="" type="checkbox"/> Specials excluded
<input type="radio"/> Summary only	<input checked="" type="checkbox"/> Output records/file	<input checked="" type="checkbox"/> Output * DATE	<input checked="" type="checkbox"/> Others excluded
Abort if findings > <input type="text" value="0"/>		<input checked="" type="checkbox"/> Suppress text	<input checked="" type="checkbox"/> Tag_ASSO -> ASSO
		<input type="checkbox"/> List files without finding	

GedTagSearch Definition File

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

? New Info **Start** Edit Log Close

© Diedrich Hesmer, 2015-2024 Version: 2.4.0

File "C:\DatVS\VS 2015\Projects\iniFiles\ExclFileGts.txt" read

Functionality

This program scans ged-files according defined tag-text-patterns. INDI, FAM, SOUR and/or _LOC records will be scanned. Depending on the option, detail data such as the found names/families with or without the tags with their texts found, the line numbers and the number of findings will be listed. The scan can run up to 4 tag levels. You can also search for

- Tags without text
- Missing tags
- Multiple tags

A log file is created during processing, which can be opened with any text editor.

Description

The description is divided into the program sequence, the search criteria, the options, the buttons and their functions and the log-file.



Program sequence

After starting the program above screen will show up.

There are 3 tabs for definition of the selections, for listing of the files found and their selection for analysis and for listing of the findings.

Within one run up to 6 Gedcom tag combinations in any files may be searched for. For this a "Start-folder", at least 1 file (also ".") and 1 tag combination must be defined. Beside the Start-folder also several sub-folders and from those its sub-folder may be scanned.

- At the first call all fields are empty resp. set to the system values.
- Definition of Start folder and files.
- The selection criteria for the ged-files to be analyzed and the tag-text pattern must be defined.
- Selection of the options.
- By "Start" the processing starts. A search of ged-files will start based on the defined start folder, the sub-folders and the ged-file definitions. Files found will be listed in tab "List of files" where they may be checked and selected for further analysis.
- By "Search" the checking of the files selected according the defined tag-text pattern will start.
- The findings for each file will be listed in the "List of results" and the log-file according the output options.
 - Line number, record type, tag combination, record nbr, i.a. name of person, i.a. lines of further findings
 - Number of combinations found
 - Found tags and numbers for "1 *" tags
 - Total combinations found.
 - List of files with no findings
- The log-file with the file name "GedTagSearch.log" is stored in the start-folder.
- During a next start of the program, it will search for an ini-file. If found, the settings of the last "Close" will be used and the fields filled accordingly. They may be modified before pushing "Start".

GedTagSearch Definition File

The "ugts" definition file contains all settings of the main screen for later re-use.

- By the button [Symbol with a diskette] the storage location and the name of the file can be defined and saved as ugts file.
- By the button [Symbol with open folder] a stored ugts file can be loaded and the settings can be transferred to the screen.

Start folder and Files

Starting from a defined start folder any sub-folder and ged-files to be scanned may be defined.

- For folders the following applies
 - Only the Start-folder, optional with all its Sub-folders should be scanned or
 - The Start-folder with additional defined Sub-folders, optional with all its sub-folders should be scanned.
 - In case of multiple defined Sub-folders they must be separated by ";".
Example: By "valid;conv" the Sub-folders "valid" and "conv" of the Start-folder will be scanned.
- For the selection of files following applies:
 - An entry must be entered here.
 - The file names have to be entered without the extension ".ged".
 - Multiple file names must be separated by ";".
 - Within the file names the wildcard characters * # ? may be used.
Example: "*_u;*_sor" will process ged-files with file names ending by "_u.ged" and "_sor.ged".
 - If only the wildcard characters "*" is entered, all ged-files will be analyzed.
 - ged-files created by GSP-, Special- or Other-programs can be de-selected. See next para.



Control file "ExclFileGts.txt"

This is an optional file. It is used to exclude ged files whose filename contains certain "significant characters" from the analysis. These are identified by the last 3-5 characters (so-called significant characters) in the file name of the ged file, i.e. the characters in front of ".ged". Example: The ged file has the name xxx.ged. The GSP reducer turns it into a xxx_red.ged, the GSP sorter into a xxx_sor.ged. Thus, the significant characters are "_red" or "_sor". The contents of this control file replaces the system values. The file is a single-line text file that must be saved in the ini file directory. It contains the last significant characters from the names of the ged files to be excluded.

Structure of the line:

- It contains 3 groups, separated by semicolon ";": **GSP_group**;**Specials_group**;**Other_group**. Thus 2 semicolons are required. If a group is not defined, the system values are taken for it.
- Each group contains the significant characters of the ged files to be excluded, separated by comma ",".
- Example file (in one line):
`_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`

Of course, any other text that corresponds to the end of a name of the ged file can be defined as significant characters.

Search criteria

These defines the scan for the required combinations of tag-text pattern. Incorrect entries are highlighted in pink and must be corrected.

Up to 4 different combinations of tags with or without text can be defined. Used rules:

- Selection of the records to be searched: INDI, FAM, SOUR, _LOC. Up to 2 of these record types can be selected for every definition line.
- Definition of the level-1 tags (TAG_1) in the format "1 Tagname" resp., if the tag is followed by text, "1 Tagname Text". As special case "1 *" is allowed to scan all level-1 tags. Optional the level-1 tags found and their numbers may be listed.

Special feature: In the special case "1 *", level 1 tags that should NOT be analyzed can be deselected. For this purpose, the tag names after the "1 *" must be specified by semicolon ";" and tag name. Several tag names must be separated by semicolon ";". Example: "1 *;BIRT;DEAT;MARR".

- Definition of the level-2 (TAG_2), level-3 tags (TAG_3) and level-4 tags (TAG_4) in the format "x Tagname" resp., if the tag is followed by text, "x Tagname Text". These fields are optional and usually filled with a text following the tag.

Note: If the level-2 tag is a TYPE tag, then the field for TAG_3 must be a level-2 tag or empty, TAG_4 must also be empty.

- The "Tagname" must represent the complete tag name. All "Text" of level-1 to 4 may include the wildcard character * ? #. In addition, a text pattern comparison is allowed as described at GSP User Guide part 1, at appendix "String Comparison". The comparison of the text is case sensitive.

Note: To search for * # ?, they must be enclosed by parentheses [...], e.g. "[?]" . These fields may be empty, if the HEAD options are selected.

Special control characters:

- **Missing Text:** If a single "%" character is entered as "Text" for levels 1-4, the corresponding tag without any text is searched for. Example: "1 SEX %".
- **Missing Tag:** If a single "-" character is entered as "Text" for levels 1-4, those records are searched for, where the corresponding tag is missing.

Note: "1 * -" is not allowed, only "1 -" or "1 +" may occur in each line.



GSP Utility – GedTagSearch

Record		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 -			

In the above example, line

- 1 searches INDI records for multiple DEAT tags
 - 2 searches INDI.BIRT for missing DATE tags
 - 3 searches INDI and FAM for missing _GOV tags subordinated to PLAC subordinated to all level 1 tags, except MARR and CHR
 - 4 searches INDI and FAM for all lvl-1 tags for all PLAC.MAP tags those LONG texts starting with "W-". The wildcard character "*" is important so that the "-" is not recognized as a control character.
 - 5 searches _LOC for missing _GOV tags.
- **Multiple Tag:** If a single "+" character is entered as "Text" for levels 1-3, those records are searched for, where the corresponding tag exists multiple times. **Note:** Only 1 "-" or "+" may occur in each line.

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

In the above example, the program searches in line

- 1 in INDI and FAM records for multiple level-1 tags except for OCCU and OBJE
- 2 in FAM for multiple MARR tags
- 3 in INDI for missing _LOC tags under DEAT.PLAC
- 4 in INDI for multiple EVEN tags with all TYPE texts.

A findind is given if a match with the defined criteria was found in the ged file.

Options

Following options are available:

- As alternate Options:
 - **Output quantity found** – the number of tag combinations found in every file and in total. This also will be listed at *line numbers*.
 - **Output line numbers** found with the tag combinations -> line-nbr., tag combination, record number. i.a. name of the person.
 - **Summary only** lists only the total values.
- Output of the name of the genealogy program (**HEAD.SOUR**) and Character set (**HEAD.CHAR**) for each ged-file.
- **Output records/file**
- **Output tags found** lists the tag combinations with their texts.
- **Output "1 *" tags** lists the tags found and their numbers for every file and in total.
- **Output * DATE** lists the birth resp. christening date after the name.
- **Suppress text** lists the tag names without their text.
- **List files without finding** writes filename and path to the log File.
- **GSP / Specials / Others excluded** will exclude these types of files.
- **Tag_ASSO -> ASSO** converts during reading of the file a tag "_ASSO" to "ASSO".
- **Abort if findings > nnn** causes an abort for more findings than defined. A value of "0" skips checking."



Buttons

In addition to the ughts control file buttons described above, following buttons exists:

- [Symbol with an open folder] Defines the start folder for searching ged-files.
- [x] right side of the lines deletes all entries of the line.
- [x] below of the tag definitions deletes all entries of the tag definitions.
- [^] allows to move one of the 5 lines (2..6) by 1 position upwards. To do this, mark the TAG_1 text box in the line to be moved and then press this button. The content of this line will be exchanged with the line above.
- [?] Opens the description.
- [New] Clears all definitions.
- [Info] Shows the folder where the ini-file is stored (contains all settings at the last "Close").
- [Start] Starts the processing.
- [Edit Log] Opens the log-file by the text editor.
- [Close] Writes the ini-file with all settings and closes the program. Closing the program by the little "x" at the upper right corner of the screen will not write the ini-file.)
- [Search] at tab "List of files" starts the scan.

Examples

They are found at tab " Examples".

Extract of a Log-file

The file lists a summary of the analysis and i.a. for each finding:

- Line number and search structure.
- Record Id.
- Name of the person of the respective record.
- Actual tags with attribute text, for multiple tags the additional line numbers.
- Summary of the findings incl. list of files with no finding.

Here an excerpt of a file:

```
05.11.2017 18:52:32 Start Scan - GedTagSearch version: 1.2.0
Line-Nbr: Tag combination -> Record-Id
*** File: "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 found in 1 file(s):
    1 x FAM: 1 *.2 HUSB.3 AGE *
    1 x ENGA
    1 x IND: 1 ADOP.2 AGE
    0 x IND: 1 SEX U
Total records - INDI: 23, FAM: 10

No findings in 2 file(s):
    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 18:52:32 Analysis of files completed
```

Versions

(N = New, C = corrected, M = modified)

- 1.0.0 – 25.02.2017 – N: Initial publication
- 1.0.1 – 01.07.2017 – M: License function
- 1.0.2 – 01.08.2017 – N: Limit for abortion
- 1.1.0 – 10.10.2017 – N: Output of "*" 1" tags found



GSP Utility – GedTagSearch

- 1.2.0 – 20.11.2017 – N: Checklist for files, list of files w/o findings, lvl-2 TYPE
- 1.2.1 – 19.02.2018 – M: Checking existence of text level 2/3
- 1.2.2 – 08.05.2018 – N: "?" Show Docu
- 1.3.0 – 03.03.2020 – N: Control file "ExclFileGts"
- 1.4.0 – 31.03.2020 – N: Option "Suppress text"
- 1.5.0 – 30.10.2020 – N: "%" character as Text lists tags without any text
- 1.6.0 – 01.12.2020 – N: Specification of level 1 tags to be skipped at "1 *"
- 2.0.0 – 25.12.2020 – N: "-" character as Text searches for missing tags
- 2.1.0 – 20.01.2021 – N: "+" character as Text searches for multiple tags, ughts definition file, M: "°" replaced by "-"
- 2.2.0 – 01.02.2021 – N: Records SOUR and _LOC also can be analyzed.
- 2.3.0 – 01.12.2021 – N: Number of tag combinations 4 > 6, additional TAG_4
- 2.3.1 – 30.01.2022 – M: UTF-8 conversion
- 2.3.2 – 04.04.2022 – N: Output option for birth/christening date
- 2.4.0 – 06.11.2024 – N: Option "List files without finding"