TexCon

A text file conversion utility for use with Apple Macintosh computers

by

Brian D. Marshall

UNITED STATES GEOLOGICAL SURVEY

OPEN-FILE REPORT 89-165-A

Revision of Mar 27, 1989

USGS Open-File Report 89-165 consists of:

89-165-A Description and user manual (paper copy)
89-165-B Program, source code and user manual (diskette)

Brief Description:

TexCon (version 1.0) is a text file conversion program for use on Apple Macintosh microcomputers. It supports conversions between various text database, spreadsheet and other file formats. Requirements: Apple Macintosh; minimum 512K RAM. OF89-165-A, Documentation, 6 p., paper copy; OF89-165-B, Program, source code and user manual diskette (3.5 inch).

Introduction

This report describes TexCon (text file conversion) software for the Apple Macintosh personal computer, which converts various text file types from one to the other. It is meant to be a simple utility program which will save users of certain Macintosh programs from the tedium of doing file conversions by hand using a text editor or word processor. It requires a Macintosh with at least 512K and System 6.0 or later. It will run under MultiFinder with a minimum partition of 128K, although more memory will need to be allocated for converting large SYLK files.

File Formats

The file type 'TEXT' on the Macintosh refers to any ASCII representation of data. Text files are readable and easily manipulated by any text editor or word processor. TexCon works only with files of type 'TEXT'; it merely manipulates these files for convenience of data interchange.Version 1.0 of TexCon supports appropriate conversions between the following text file types: tab-text, Cricket text, QuickDEX, FactFinder, and SYLK. TexCon also has utilities for removing double quotes, removing or adding linefeed characters, removing control characters and switching commas to tabs and vice versa. TexCon may be useful with any of the following programs: Microsoft Excel, Microsoft Multiplan, MacCalc, Cricket Graph, FactFinder, and QuickDEX.

Tab-Text

Tab-Text refers to a representation of a two-dimensional array of data, such as might be generated by a spreadsheet or database program. In tab-text files, each row or record of data is separated from the next row by a carriage return, and the individual pieces (cells or fields) of data within each row are separated from each other by tab characters. Text values are not normally surrounded by double quotes, although some programs add these if the text contains a comma. TexCon limits indivudual fields to 255 characters.

Cricket text

Cricket text files are a special case of tab-text files where the first record contains an asterisk and the headers or titles for columns of data are contained in the second record. Cricket Graph reads and writes these files. Other spreadsheet programs read these files but will display the extraneous asterisk in the first row.

QuickDEX

QuickDEX files are 'TEXT' type files. Each 'card' in a QuickDEX database is stored exactly as it appears on the screen. In the file the cards are separated by an asterisk followed by a carriage return. A 'N' or a 'P' is appended to the beginning of each card to signify whether or not the card has been printed.

FactFinder

FactFinder uses binary files for its database, but it can also read and write text files, which contain all the information in the database in a certain format. A particularly useful feature of FactFinder is its ability to write a text file containing only records (FactSheets) found in the last search. It is also able to append the contents of a text file to any database. FactFinder text files are complicated because they contain keywords, dates, and names in addition to the text of the

records. FactFinder text files can be loaded into any existing FactFinder stack.

SYLK

SYLK (symbolic link) files are spreadsheet files created by many applications. SYLK files can contain many types of information; different applications read or write only the information of interest for their purposes. For TexCon's purposes, SYLK files contain an x-y grid of data. The data can be numeric or text. TexCon only looks at the values of the cells in a SYLK file; it does not check the formatting. This means, for instance, that numbers formatted as money or dates are read only as pure numbers. For plotting purposes, pure numbers are usually preferable. If the formatted data (as displayed) is desired then saving the spreadsheet as tab-text (if the program does this) may give the desired result.

Text Database Conversions

Text database programs are databases that store only text with minimal formatting. They have key-word or full character searching capabilities.

FactFinder to Text

This conversion takes a FactFinder text file and removes the name, dates, and keywords as well as the backslash (\) delimiters. The resulting text file contains only the factsheet text records separated by carriage returns. If a factsheet record contains internal carriage returns they will still be present in the text file. This conversion is useful for importing text records found by a keyword search in FactFinder into a word processor without having to manually remove all the extraneous information in the FactFinder text file. If the file chosen for conversion does not contain the correct header indicating that it is a FactFinder text file, the conversion will stop with a sound to indicate there was an error.

Text to FactFinder

This conversion takes a text file in which individual records are separated by carriage returns and makes a FactFinder text file containing these individual records as factsheets. The name for each factsheet is contructed from the first 25 characters of the text record or until a carriage return is encountered. The created and modified dates are set to the current date returned by the Macintosh clock. 'FROM TEXT' is added as the only key word identifying the records. Additional key words can be added by using the automatic key word capability of FactFinder prior to loading the text file. Extra carriage returns between records will *not* create extra blank factsheets.

FactFinder to QuickDEX

This conversion takes a FactFinder text file and converts the factsheets to a new QuickDEX file, making each factsheet a card in the new file. All characters are passed to the card except asterisks, which QuickDEX uses as delimiters. If a factsheet contains more than 512 characters, those beyond the 512th character will not be written to the new file since QuickDEX cards are limited to 512 characters. If the file chosen for conversion does not contain the correct header indicating that it is a FactFinder text file, the conversion will stop with a sound to indicate there was an error. The QuickDEX file that is created will not be recognized by QuickDEX unless the option key is held down while selecting Open from the QuickDEX menu.

QuickDEX to FactFinder

This conversion takes a QuickDEX file and converts it to a FactFinder text file, which can then be loaded into any FactFinder database. Each card becomes a factsheet. The name for each factsheet is contructed from the first 25 characters of the text record or until a carriage return or asterisk is encountered. The created and modified dates are set to the current date returned by the Macintosh clock. 'FROM QUICKDEX' is added as the only key word identifying the records. Additional key words can be added by using the automatic key word capability of FactFinder prior to loading the text file. If the file chosen for conversion does not contain the correct header indicating that it is a QuickDEX file, the conversion will stop with a sound to indicate there was an error.

Spreadsheet Conversions

Spreadsheet programs are probably familiar to almost all personal computer users. They are flexible programs that store and manipulate limited amounts of data. Their basic feature is an array of cells in which individual data items are stored.

SYLK to Text

This conversion is by far the most complicated and memory hungry one attempted by TexCon. The reason for this is that the data in SYLK files is not necessarily written in sequence; data cells near the origin of the spreadsheet may be encountered after cells farther from the origin. Therefore, a data structure is created in memory as the SYLK file is read so that the data in memory will be in correct order for creating the text file. Successful conversion of a SYLK file requires enough available memory to hold the entire array of cells. When selected, the SYLK to Text conversion will display the number of cells that can be stored in your computers memory; additional memory can be made available by turning off the RAM Cache, disabling INIT utilities, or allocating more memory to TexCon (if running MultiFinder). However, the lack of enough memory will only cause a partial conversion; when no more cells will fit, the computer will beep to indicate an error and the user is given the opportunity to save the partially converted file. The SYLK file is converted to a tab-text file in which rows of the spreadsheet are separated by carriage returns and cells within rows are separated by tabs. Text strings within cells are not double-quoted. If the file chosen for conversion does not contain the correct header indicating that it is a SYLK file, the conversion will stop with a sound to indicate there was an error.

Text to SYLK

This routine reads a tab-text file and creates a SYLK file. Leading and trailing double quotes are stripped before creating the SYLK records. Individual data items are limited to 255 characters; any characters beyond 255 will be ignored.

SYLK to Cricket Text

This routine is identical to SYLK to Text but creates a text file that Cricket Graph will recognize as containing column headers created from the first row of the spreadsheet. Cricket Graph will also recognize which columns contain numeric rather than text type data and format the columns accordingly. This utility makes it easy to take any spreadsheet data and plot them using Cricket Graph.

Cricket Text to Text

This routine simply removes the leading asterisk and carriage return used by Cricket Graph to indicate that the first row contains the column headers. If the file chosen for conversion does

not contain the correct header indicating that it is a Cricket Text file then the conversion will stop with a sound to indicate there was an error.

Text to Cricket Text
Some applications (e.g. Microsoft File) create text files with the field names as the first record, but will not read into Cricket Graph correctly. This routine simply adds an asterisk and a carriage return to the beginning of any text file so that Cricket Graph will recognize the first row as column headers. Leading and trailing double quotes are stripped before creating the text records.

Other Conversions

These conversions either remove or add characters from/to a text file or exchange one character for another.

Remove Double Quotes
Some applications put double quotes (" ") around text strings that contain commas. Since other applications may not strip these quotes off when loading the file, it can be useful to have this utility which strips all double quotes from a file.

Remove LineFeeds
Linefeed characters are used on many personal computers (including IBM compatibles) in addition to carriage returns to signal the end of a line. Since these characters may cause problems when a text file written on one of these machines is read on a Macintosh, this utility strips these characters from the file.

Add LineFeeds
Linefeed characters are added to the file wherever there is a carriage return, thus creating <CR><LF> sequences required to read the file on some other personal computers.

Commas to Tabs
Some applications may use commas as delimiters in spreadsheet text files. Since most standard Macintosh applications will not read such a file, this utility changes all commas in a file to tab characters. Commas occurring within quoted strings will not be changed. Tab-delimited text files are a standard format on the Macintosh.

Tabs to Commas
This routine does just the opposite of Commas to Tabs in case one wants to go the other way. Tabs occurring within quoted strings will not be changed.

Remove Controls
This option removes all non-printing ASCII characters and extended ASCII characters. In other words it takes a file and creates a new one with only the following characters:

Character	ASCII number
Horizontal Tab	9
Carriage Return	13
Space	32
!	33
"	34
#	35

$	36
%	37
&	38
'	39
(	40
)	41
*	42
+	43
,	44
-	45
.	46
/	47
0 thru 9	48 thru 57
:	58
;	59
<	60
=	61
>	62
?	63
@	64
A thru Z	65 thru 90
[	91
\	92
]	93
^	94
_	95
`	96
a thru z	97 thru 122
{	123
|	124
}	125
~	126

This routine can be useful if a word processing file can't be read because the file is damaged or because you do not possess the appropriate application program. Using ResEdit or a similar file-editing utility, you can change the file type of the bad file to 'TEXT' and then use this TexCon routine to strip out most of the control character sequences. The resulting file can then be opened using a word processor or text editor. The formatting will be partially lost, but the text should be intact. Depending upon the damage to the file, there may still be unwanted characters that you will have to remove. I think that this beats re-typing in most cases.