

TextBroom3

TextBroom 3 by Blue Line Studios

Copyright ©1999 - 2001 Blue Line Studios. All rights reserved.

Software programming, interface design, manual written by Daniel C. Kueng.

Edition February, 2001

This manual and the software described are copyrighted with all rights reserved. The manual may not be distributed, sold, or rented to any other person than the owner of the license, wether in whole or part, without prior written consent of Blue Line Studios. Mentioned product and company names may be registered trademarks. Use of the software described is at the user's sole risk.

Blue Line Studios

CH 4000 Basel/Switzerland

Web: <http://www.blueline-studios.com>

E-mail: studio@blueline-studios.com

INTRODUCTION

What this application is for

TextBroom 3 helps you processing text files. Perhaps you just exported a spreadsheet from Excel as "plain text", and the resulting file contains odd numbers of tabulators; e. g. 10, 12, 17, and 35 in a row. You would have to use a search & replace function, perhaps the one found in QuarkXPress, search for 35 tabs, replace with 1, next, search for 34 tabs, replace with 1, and so on.

Or you got a text file someone has written who didn't know about the common typesetting rules. The file would contain spaces instead of tabulators to indent lines, again odd numbers of them; it would perhaps miss spaces after commas and periods; it would contain line breaks instead of continuously written text; it would contain strange number separators. In short, it would be the nightmare of every typesetter. The accompanying printout from MS Word would be looking great though.

TextBroom 3 comes to the rescue, letting you fix these often-made typos with just one click. Some of the functions are even highly customizable.

Important note: TextBroom 3 works only on plain ASCII (text-only) files.

HOW TO USE TEXTBROOM 3

Getting started

Before you try to open a text file, be aware that TextBroom 3 can only work with ASCII text-only files; e. g. if you have a native MS Word file you should open it with MS Word, and save the file (or a copy of it) as "plain text". Alternatively, run it through MacLink Plus.

Alternatively, just copy the text to the clipboard by choosing *Copy*, then switch to TextBroom and paste it into its text field.

TextBroom's default memory allocation is pre-set to 22 MB for the PPC-version and to 18 MB for the 68K-version. If you work with large files (10 MB and more), it is recommended to increase the memory allocation to 3 times the amount of the files' size you want to work with. In the Finder, select TextBroom, then choose menu *File: Get Info*, or press the keyboard shortcut *Command-i* to display the Finder's info window. Type the new memory allocation into the field *Preferred Size*.

Loading text files

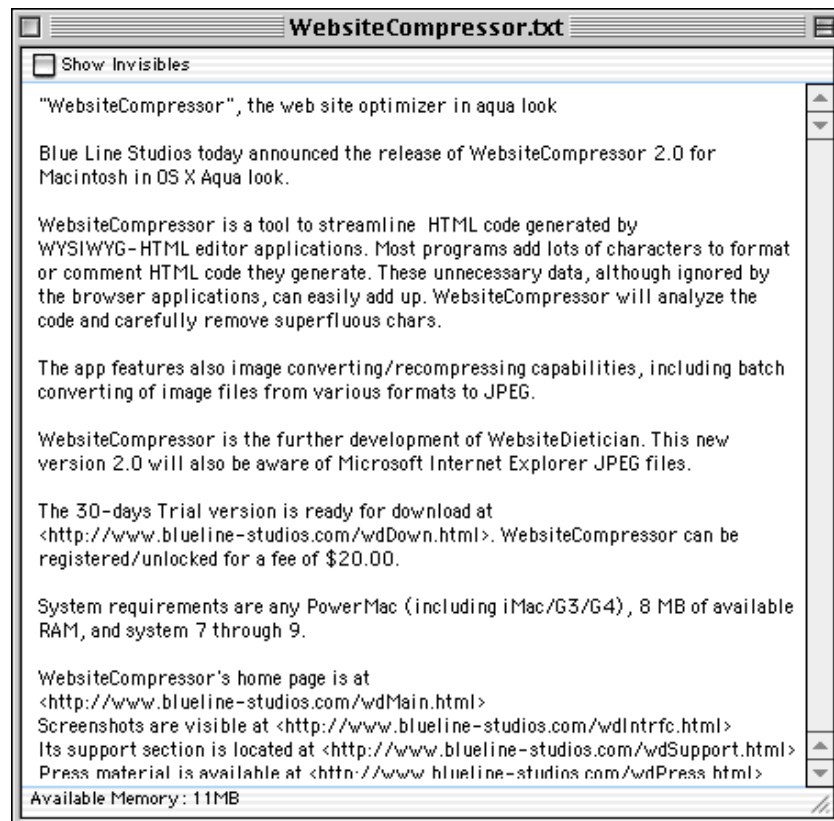
When launching TextBroom, you will be presented with the main window displaying the "Multiple spaces" screen. This function will be explained later.

Choose the menu *File: Open Text File...* or press the button *Open Text File...* Alternatively, use the keyboard short cut *Command-O*. If your Mac OS supports Drag & Drop, you can instead just drop the file from the Finder into TextBroom's window.

Note: If the file exceeds 32 KB, only the first 32000 characters can be displayed in the text editing window; the entire file, however, will be loaded into the RAM and will be processed . For tech reasons, the text cannot be manually edited.

TextBroom 3 will let you select files of type "TEXT" by default. However, there is a backdoor: hold down the *Option* key while clicking the *Open Text File...* button (or the menu equivalent), and TextBroom 3 will allow you to read just about **everything**.

ATTENTION: Be careful with this feature, as you might easily crash your system, especially when opening files that are frequently used by the operating system.



Click the titles on the black panel at the left side of the window or choose menu *Task* to jump to the appropriate section.

THE FUNCTIONS EXPLAINED

Multiple spaces

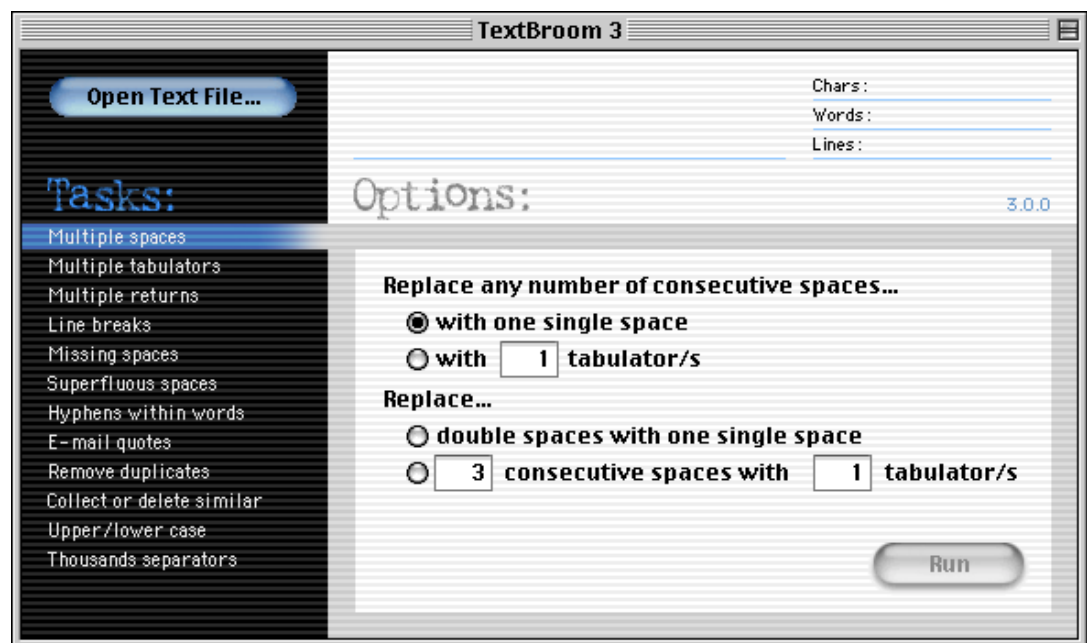
It is common practice to type spaces to indent lines, instead of doing it the proper way using one tabulator. Although that MS Word printout might look great, it causes a mild frustration to every typesetter trying to use the text in a layout application. As these spaces appear more often than not in odd numbers, it's not possible to just search 31 spaces & replace with 1 tab (See "Replace multiple tabulators").

Radio button #1: All multiple spaces in a row (odd numbers) will become 1 single space.

Radio button #2: Specify the number of tabs all spaces in a row (odd numbers) will yield to.

Radio button #3: Specify how many spaces in a row will yield to your specified number of tabulators.

If you want to replace any number of consecutive spaces, or a certain number of them, it's recommended to check *Double spaces to single space*. As double-spaces are an often-made typo, 2 consecutive spaces will otherwise yield to 1 tab.



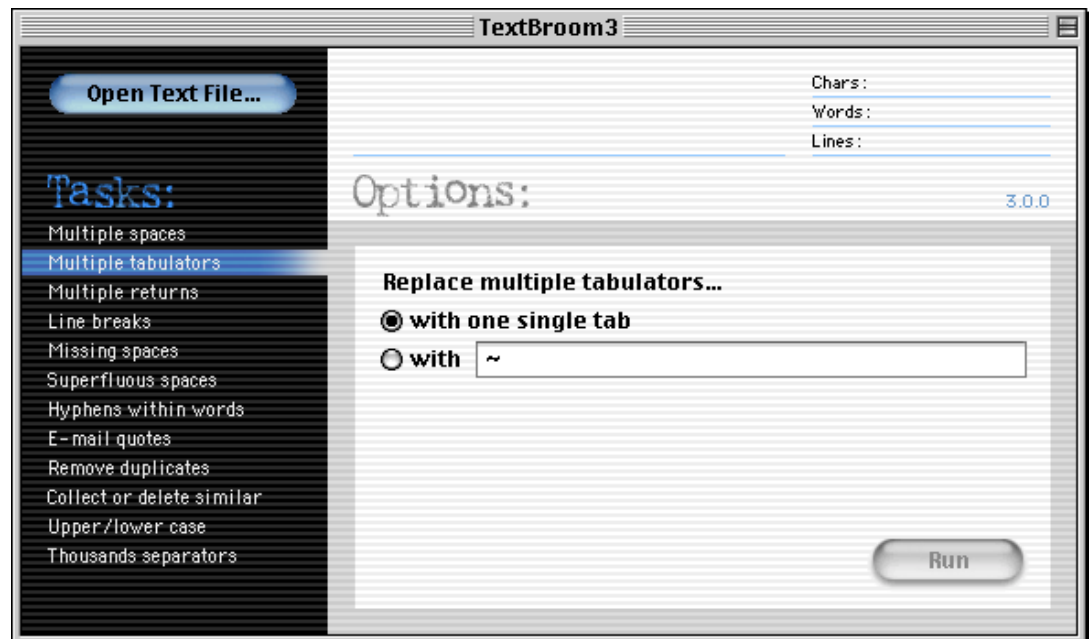
Note; All running processes can be aborted by either pressing the *Stop* button or by typing *Command-period*.

Multiple tabulators

A known problem when exporting plain text from Excel spreadsheets is that the resulting text file will contain lots of consecutive tabulators. A simple searching for tabulators and replacing with nothing won't do, because you want to retain at least 1 tab, defining each column.

The hard part is that these consecutive tabulators appear in odd numbers; e. g. 20 tabs, 28 tabs, 29, 37, and 41. Using a program such as QuarkXPress, you would have to search for 41 tabs, replace with 1, "Find first", search for 40 tabs, replace with 1, and so on.

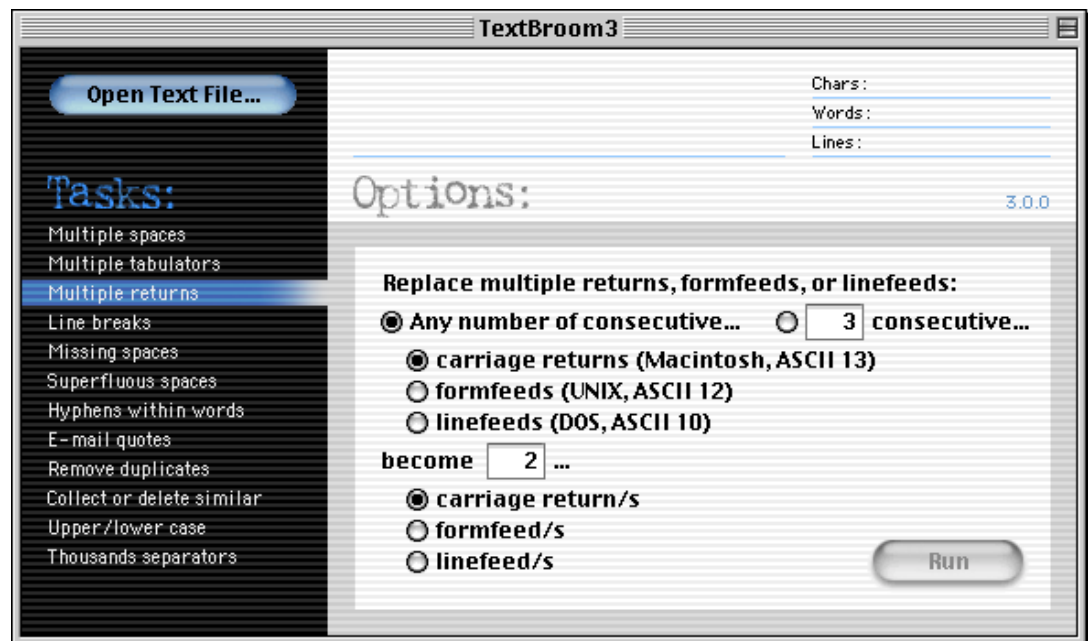
Instead of replacing odd numbers of consecutive tabulators with one single tab, you can put in your own string or symbol as a temporary placeholder, which can later be easily replaced within your favorite text editor or layout application.



Multiple returns Multiple consecutive returns in texts are only a mild annoyance to typesetters. All the same it's more professional to get rid of them.

Radio button #1: Specify the number of consecutive returns (carriage returns; ASCII 13) that you want to become the specified number of returns.

Radio button #2: Any number of consecutive returns will become the number of returns you specify.

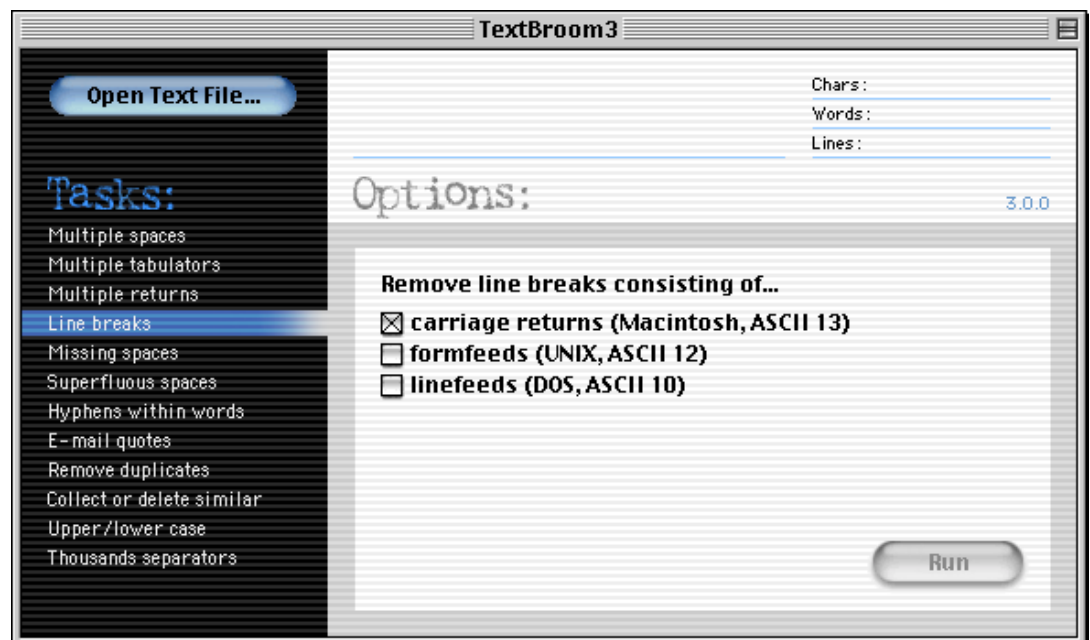


Line breaks

Some typists type a return after each line, although the text wraps automatically in just about every word processor. E-mail clients also tend to hard-wrap text before transmitting the data.

Searching for carriage returns and replacing with nothing won't do, since legitimate double returns would be reduced to a single return. You would later have to search for single returns and replace with 2 returns to restore them.

Also, some texts that came from the DOS platform might not even contain carriage returns (ASCII 13), but linefeed or formfeed characters.



Missing spaces

Sometimes, text misses space characters, e. g. after commas and periods.

However, you wouldn't want to mess up periods used as a decimal separators within numbers, or commas used as thousands separators. By default, TextBroom 3 will check if the word containing the comma or period is a number.

Checkbox #1: Puts a space after every comma character if space is missing.

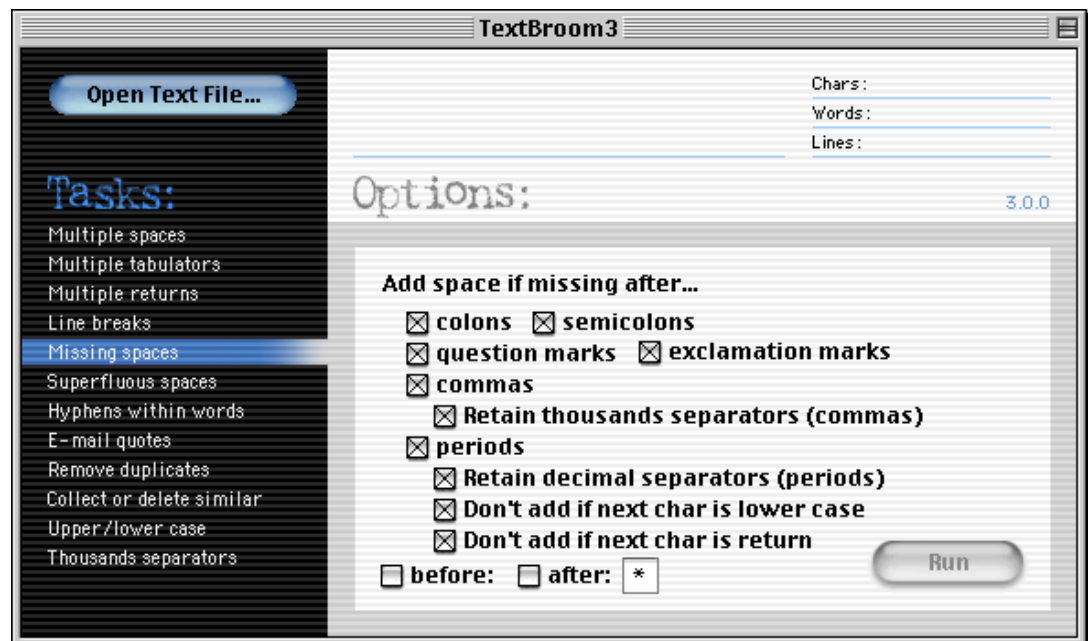
Checkbox #2: Puts a space after every period character if space is missing.

Checkbox #3: Ignores periods within numbers.

Checkbox #4: Ignores commas within numbers.

Checkbox #5: Ignores commas and periods if the character following the comma or period is a Macintosh carriage return (ASCII 13).

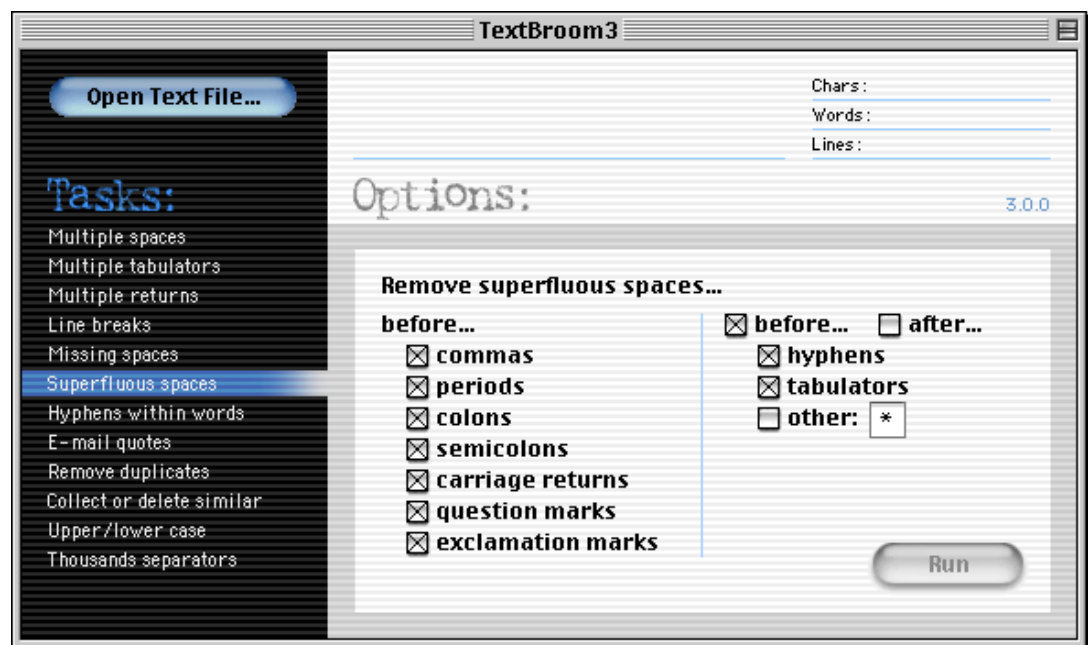
Checkbox #6: Lets you specify your own search character.



Superfluous spaces

Typesetters stick to certain rules chiseled in stone: never should spaces be preceding commas, periods, question marks, or exclamation marks, as well as a number of other characters.

Same is true for tabs, as spaces preceding and following tabulators might not position the indents in your layout program accurately. Carriage returns are better not preceded by a space character either. However, keep in mind that the space character before the return may very well be needed if the text is manually wrapped. **If you also need to remove line breaks, run the *Line breaks* filter previously, or uncheck the box *carriage returns*.**

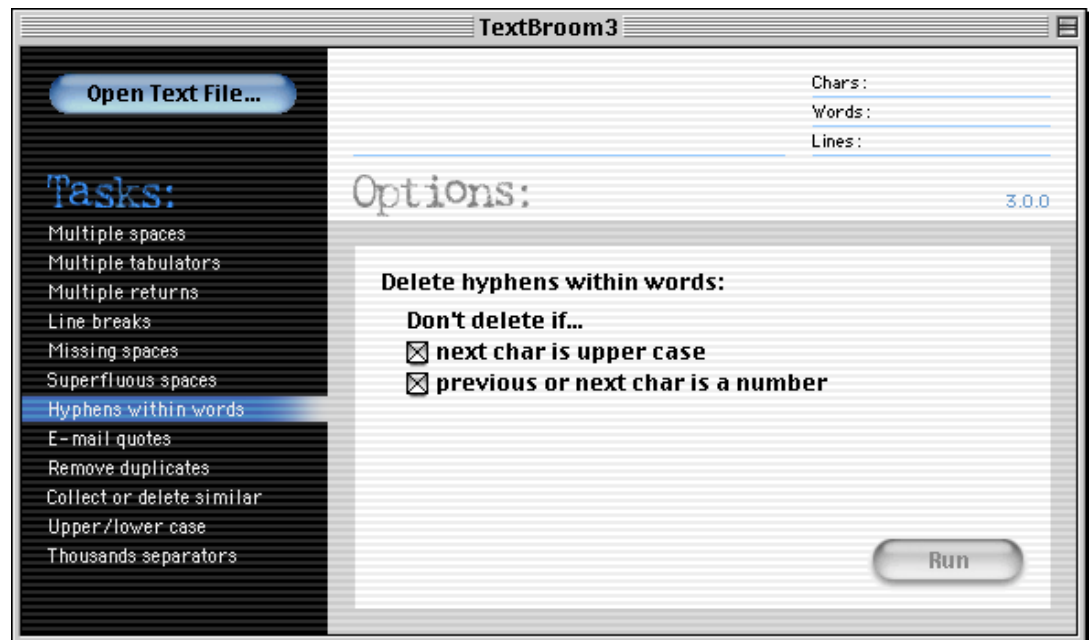


Hyphens within words

After having removed line breaks, some words might now contain superfluous hyphens, as hyphen-return might have been typed to break words manually. However, sometimes hyphens within words are legitimate; it could be a minus sign in front of a number, a prefix in a phone number, or a name using hyphenation.

Checkbox #1: Ignores hyphens within numbers.

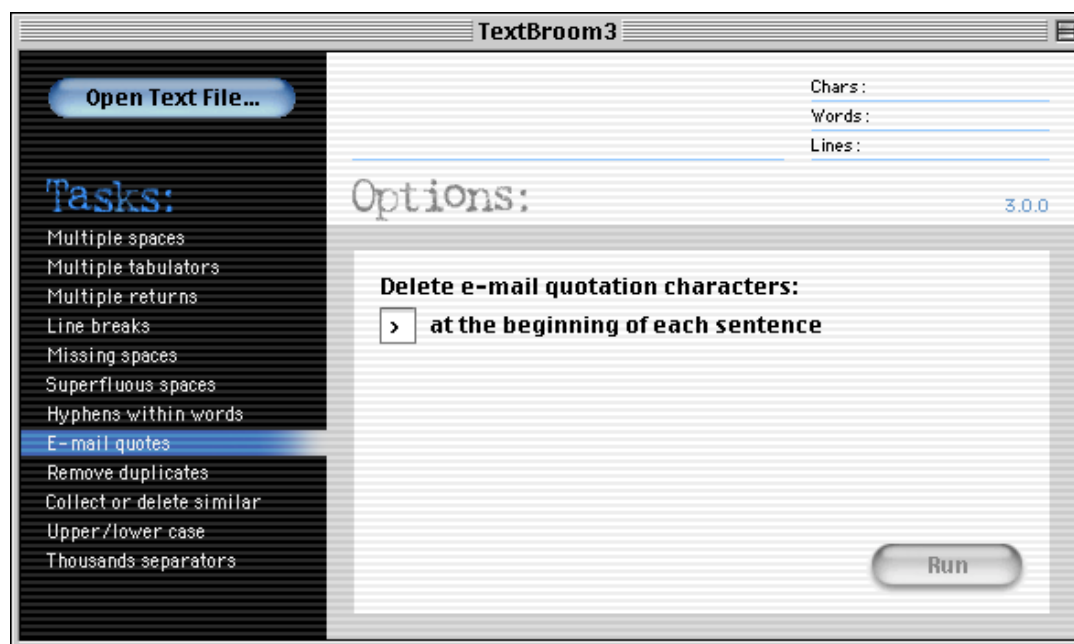
Checkbox #2: Ignores hyphens if the following character is a number.



E-mail quotes

Perhaps someone e-mailed you a text, which has been e-mailed to and fro half a dozen times and now at the beginning of each line there are six e-mail quotes in a row.

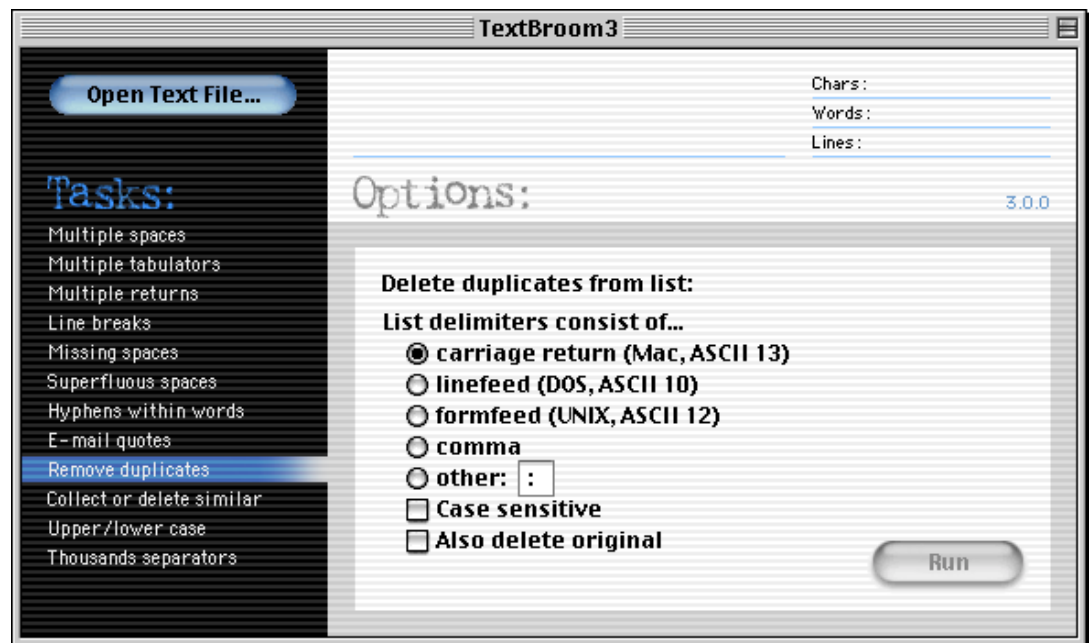
In that case, you will need to press the button six times in a row.



Remove Duplicates

Assume Mr Sisyphus got a list of addresses and he suspects it to contain multiple entries. If the list is extensive, this function comes in handy.

Specify the chunk delimiter, e. g. *carriage return* if you want to look for lines delimited by Macintosh returns, or *comma* to search a comma-delimited database. *Other* will let you specify any other delimiter, including space.

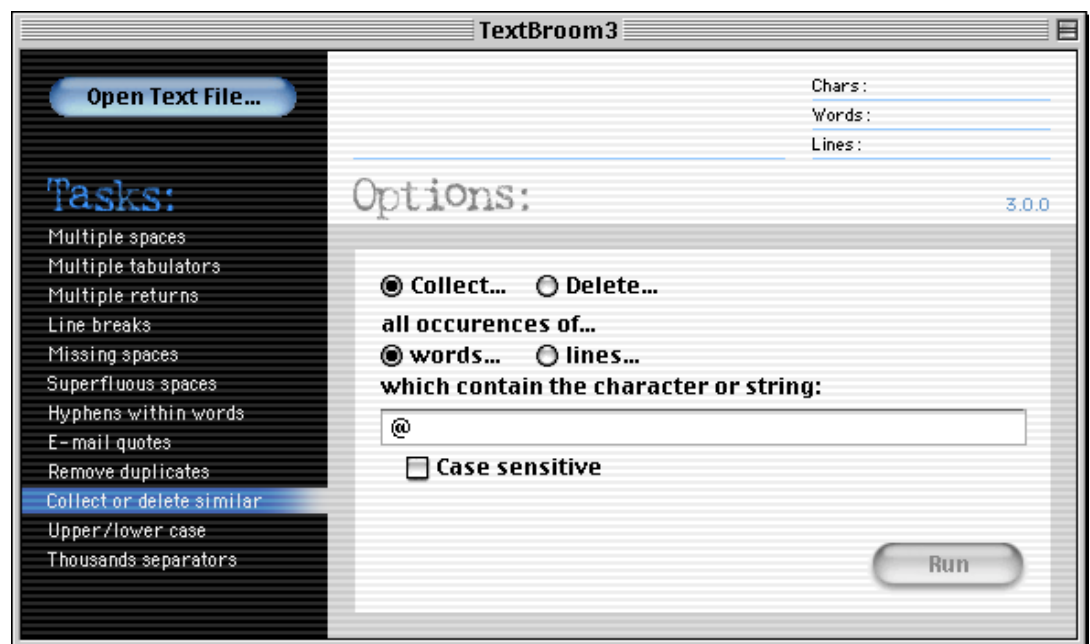


Also delete original will delete **all** lines that appear more than once, meaning every chunk of which a duplicate exists will also be removed.

Collect or delete similar

This function is great if you need to collect e-mail addresses which are littered within a huge text file or database. As every e-mail address must contain the "@"-symbol, simply put this character into the search string field and TextBroom 3 will collect either all lines (Macintosh carriage return-delimited, ASCII 13) which contain e-mail addresses, or single words; the addresses.

This will also work when searching for strings, e. g. "www." will find all World Wide Web URLs buried in a text.



Upper/lower case

Here are some basic upper/lower case functions, as well as some smart ones:

Radio button #1: Will change all text to upper case.

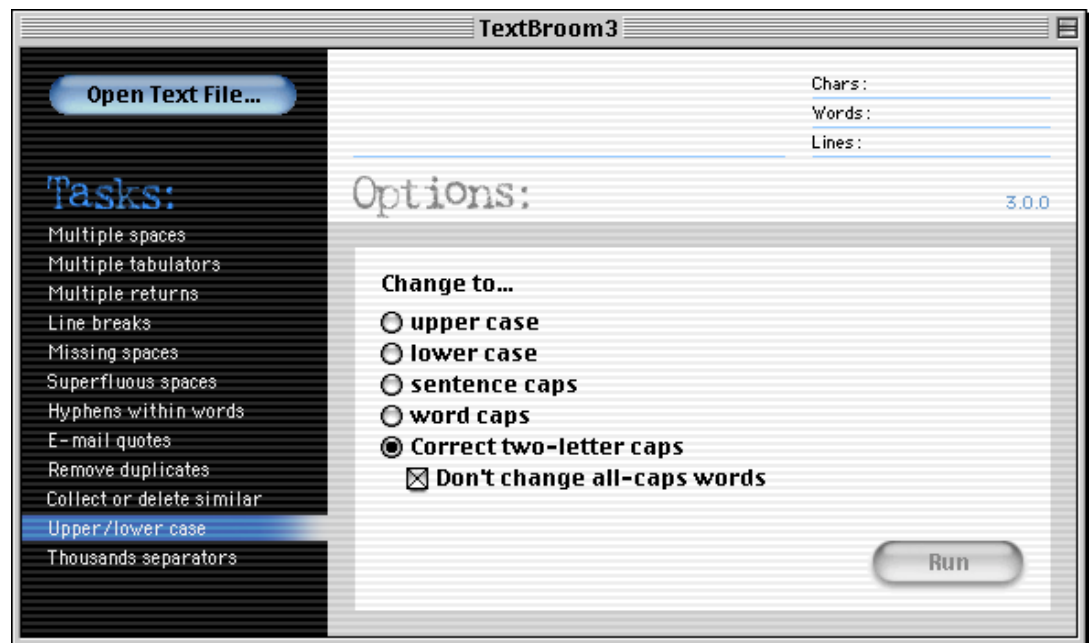
Radio button #2: Will change all text to lower case.

Radio button #3: Only the first character of each sentence will be changed to upper case.

Radio button #4: Only the first character of each word will be changed to upper case.

Radio button #5: A common typo is to hold down the *shift* key for a fraction of a second too long, resulting in two consecutive capitals.

Checkbox #1: *Correct two-letter caps* will **not** change words that are **all** upper case.



Thousands separators

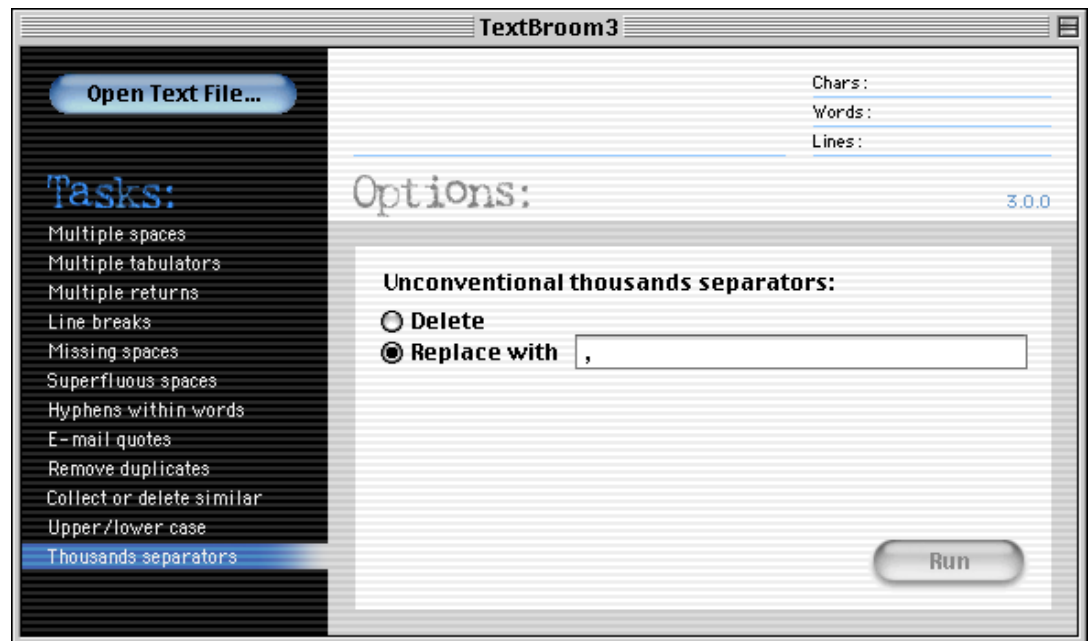
The ultimate typesetter's nightmare: imagine an abysmal text with numbers which contain the wrong thousands separators. How would you, or rather, the search & replace function in your layout app, distinguish between legitimate commas and spaces in the text, thus only picking the numbers?

Take the string "12'000" for example, containing an apostrophe; in the English language, thousands are separated by commas. Impossible to search for apostrophe and replace with comma, because that would change words like "it's" to "it,s". Even worse, imagine the thousands separator to be a space character..

This function will only change number strings, never text strings.

Radio button #1: Delete **any** thousands separators.

Radio button #2: Put the character specified into all thousands separators. This option lets you later replace your unique string (e. g. §numSep§) in your layout app with a 1/4 space.



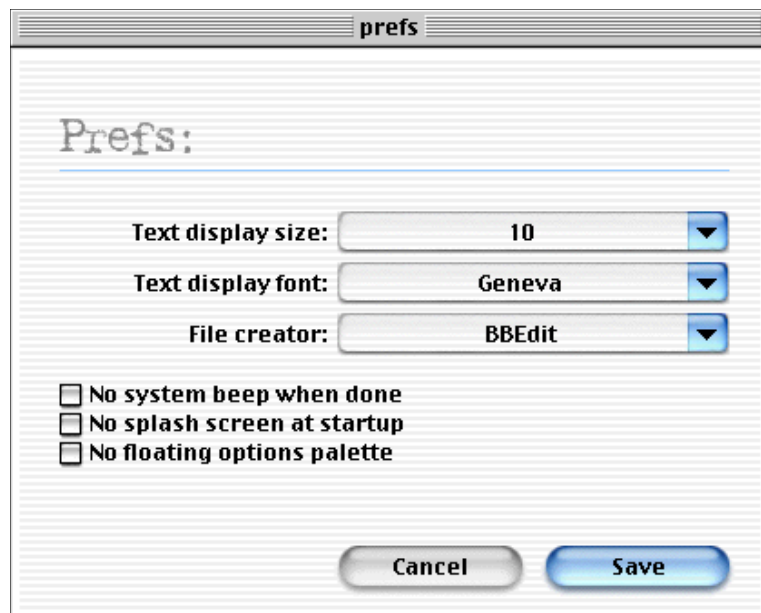
Setting your preferences

Selecting menu *Edit: Settings* will display the Preferences panel. Alternatively, press the keyboard equivalent *Command-K*.

Here you can set the file creator code of the files TextBroom 3 writes out, set the text window's text size, specify the font, as well as enable the splash screen to display on startup and turn off the system beep at the end of the processes.

Popup button *File Creator: Other...*: Set your own creator code* for the files TextBroom will write by selecting any document or application of your choice. The button name will change to the selected creator code.

Note: When saving files, TextBroom 3 will retain the original creator type, regardless of these settings. **Only when using the *Save As...* command, will the new file have the creator signature of your choice.**



*All Mac files have an 8-letter creator/type signature assigned: Communicator files for example have the creator code "MOSS", type code "TEXT". The application Communicator itself is therefore "MOSSAPPL", whereas type "APPL" defines it as executable code; SimpleText creates "txtTEXT" documents, Explorer "MSIETEXT", etc.

TextBroom can only handle files of type "TEXT" (plain ASCII text), the creator being irrelevant. The purpose of the "Other..." function is to provide a way to save files in the format of your favorite text editing program, e. g. BBEdit, very much as if that application had created the file. TextBroom will then write files with the signature "R*chTEXT"; it will be possible to just doubleclick that file and BBEdit will launch.

TROUBLESHOOTING

If the app doesn't work as expected

TextBroom 3 has been tested exhaustively and thoroughly. Should you nevertheless experience a problem of any sort, please make sure you follow the guidelines below.

1. Please check this list:

- Did the crash, freeze, or unexpected behaviour only happen once or does it happen every time?

If it happens everytime, please write down the exact steps that lead to the error.

- Did you restart your computer and try again?

Sometimes a restart performs wonders.

- Did you recently install any new extensions used by the operating system?

Sometimes extensions can cause problems. Please turn off the extension and try again.

- Did you trash TextBroom 3 and use a fresh backup copy?

2. If the problem persists, please take a look at our online support page at

<http://www.blueline-studios.com/tcSupport.html>

3. If you still can't find the solution of the specific problem, please contact Blue Line Studios' support department at support@blueline-studios.com. We will need an exact reproducible description of how the error occurred. Only then will we be able to track down and fix any possible flaws or bugs.

