Justification in Heirloom Troff

JUSTIFICATION
IN
HEIRLOOM TROFF

Gunnar Ritter

11/11/06

Heirloom Documentation Tools

<http://n-t-roff.github.io/heirloom/doctools.html>

Line-by-line adjustment

When determining line breaks, *troff* traditionally uses a simple method: Words are accumulated from input as long as they fit on the current output line. Once a word consumes more space than available, it is hyphenated. If a feasible breakpoint results, it is chosen; otherwise the current output line ends with the previous word.

If the adjustment mode is "b" (adjust both margins) and the line is shorter than the desired line length, interword spaces are widened to make the line fit. The line is then printed either to intermediate output or to the current diversion, if any. Afterwards, the process starts again.

This method works reasonably well but has two shortcomings: First, it is not possible to compress interword spaces by the slightest amount even if the breakpoint obtainable by widening is inacceptable. Second, if a line that fits perfectly is followed by a line with very loose spacing, it might be better to move the last word of the first line to the next one; then both lines might have less than perfect spacing, but it might nevertheless be more acceptable than one line with very wide spaces.

Both problems are addressed in *Heirloom troff*. As usual, the default behavior has not changed, though, so identical output will be produced for existing documents unless they are modified accordingly.

Shrinking interword spaces

The ".minss" request allows to specify a minimum interword space. It is only effective when adjusting both margins. It accepts an argument with the same semantics as the ".ss" request, i.e. a numeric value that is multiplied by 12/36 of the standard interword space. The space size configured with ".ss" is taken as the optimum setting. Thus with the default ".ss 12", ".minss 9" specifies that spaces may be shrunk to 75 percent.

The line breaking process is then changed such that when the first word does not fit on the current output line anymore, *troff* is allowed to shrink interword spaces to make it fit instead of deferring the word to the next line and expanding the interword spaces on the current one. *troff* has a slight preference for shrinking built-in, so if shrinking and expanding are equally far away from the optimum, shrinking is chosen. If a line can be set with the optimum setting, no shrinking is performed.

Standard adjustment settings

Harmony, liberal intercourse with all nations, are recommended by policy, humanity, and interest. But even our commercial policy should hold an equal and impartial hand; neither seeking nor granting exclusive favors or preferences; consulting the natural course of things; diffusing and diversifying by gentle means the streams of

Shrink to 67 % (.minss 8)

Harmony, liberal intercourse with all nations, are recommended by policy, humanity, and interest. But even our commercial policy should hold an equal and impartial hand; neither seeking nor granting exclusive favors or preferences; consulting the natural course of things; diffusing and diversifying by gentle means the streams of commerce, but

While the second setting is certainly not perfect, it is much better than the first one.

Paragraph-at-once adjustment

Adjusting paragraph-at-once distributes the word spaces more evenly:

Line-by-line adjustment (.ad b)	Harmony, liberal intercourse with all nations, are recom- mended by policy, humanity, and interest. But even our commercial policy should hold an equal and impartial hand; neither seeking nor granting exclusive favors or preferences; consulting the natural course of things; dif- fusing and diversifying by gentle means the streams of
Paragraph-at-once adjustment (.ad p)	Harmony, liberal intercourse with all nations, are recom- mended by policy, human- ity, and interest. But even our commercial policy should hold an equal and impar- tial hand; neither seeking nor granting exclusive favors or preferences; consulting the natural course of things; dif- fusing and diversifying by gentle means the streams of

To address the problem of an unnecessarily loose line, it is obviously necessary to look ahead to following text. Actually the best solution may involve multiple lines: the line with sufficiently tight spacing might occur several lines before the loose one, and each line in between simply starts one word earlier but contains the same number of words.

For this reason, *troff* collects the words of an entire paragraph and computes optimal breakpoints when it is ended with the next request causing a *break*. Breakpoints are considered optimal if all interword spaces in the paragraph are as close to the optimum setting as possible.

Once the optimal breakpoints have been computed, the resulting lines are output. At this time, *traps* become effective. When the entire paragraph has been printed, execution continues with the request that initially caused the *break* at the end of the paragraph.

Paragraph-at-once adjustment is enabled per paragraph with ".ad p"; the forms ".ad pc", ".ad pl", and ".ad pr" are also supported and apply the method to centered, left-adjusted, and right-adjusted text, respectively.

The request ".padj" globally enables paragraph-at-once adjustment across all environments; it is especially useful to change existing documents to use this mode.

Paragraph-at-once adjustment is compatible with almost all existing *troff* code. Most importantly, it works in combination with the *tbl, eqn, refer,* and *pic* preprocessors as well with the standard "−mm", "−ms", "−me", and "−man" macro sets.

The ".in", ".ti", and ".ll" requests should only be used to set indenting and line length for an entire paragraph. If they are used within a paragraph, breakpoints must be recomputed, and previous breakpoints are suboptimal. Documents that use such methods e.g. for inline pictures should be adapted to achieve optimum results with paragraph-at-once adjustment.

Since positions on the output line are not computed until the entire paragraph has been collected, the ".k" and ".x" number registers cannot contain meaningful values in paragraph-at-once adjustment mode. Macros that test ".k" only to determine if there is text present will work, though, since it is ensured that ".k" is never zero in this case.

The number of the current page in the "%" register can be lower than the number of the page on which the current input word will actually be printed in paragraph-at-once adjustment mode. Thus e.g. to prepare words for indexing, it is not possible to associate them with page numbers when the input is read.

Harmony, liberal intercourse with all nations, are recommended
by policy, humanity, and interest. But even our commercial pol-
icy should hold an equal and impartial hand; neither seeking
nor granting exclusive favors or preferences; consulting the nat-
ural course of things; diffusing and diversifying by gentle
means the streams of commerce, but forcing nothing; establish-
ing with powers so disposed, in order to give to trade a stable
course, to define the rights of our merchants, and to enable the

Standard adjustment settings

Harmony, liberal intercourse with all nations, are recommended
by policy, humanity, and interest. But even our commercial poli-
cy should hold an equal and impartial hand; neither seeking nor
granting exclusive favors or preferences; consulting the natural
course of things; diffusing and diversifying by gentle means the
streams of commerce, but forcing nothing; establishing with
powers so disposed, in order to give to trade a stable course, to
define the rights of our merchants, and to enable the government

Allow word spaces to be shrunk to 83 %

Harmony, liberal intercourse with all nations, are recommend-
ed by policy, humanity, and interest. But even our commercial
policy should hold an equal and impartial hand; neither seeking
nor granting exclusive favors or preferences; consulting the nat-
ural course of things; diffusing and diversifying by gentle means
the streams of commerce, but forcing nothing; establishing with
powers so disposed, in order to give to trade a stable course, to
define the rights of our merchants, and to enable the government

Adjust paragraph-at-once and allow word spaces to be shrunk to 83 %

Output-line traps have been introduced to address this issue: A "\P[xx]" in input is passed through all formatting and diversion processing along with the word it has been attached to. When the line containing it has been actually printed, the macro "xx" is executed. The behavior is then similar to a page trap. Multiple output-line traps may occur on a single line.

An index macro can use this mechanism to defer the processing of an index term until after the position of the word it refers to has been determined:

```
.nr IXcount 0 1
.de IX
.	de IX-\\n+[IXcount]
.	write index \\\\n% \\$1
\\..
\\P[IX-\\n[IXcount]]\c
..
An
.IX "index term"
index term is contained in this sample text.
```

This example macro takes the index term as a single argument. It creates a separate macro on each invocation and prepends an output-line trap calling it to the following word. The created macro then prints the current page number (processed in this macro, thus preceded by four backslashes) and the index term argument (processed in the surrounding macro, thus preceded by two backslashes).

Microtypography

To further enlarge the range available for adjustment while reducing the amount by which interword spaces are affected, *troff* also allows to vary the size of interletter spaces and the shape of glyphs with the ".letadj" request. This process is called "microtypography".

Microtypography must be applied with care. While the eye is accustomed to varying interword spaces which leave the individual words intact, varying interletter spaces and letter shapes distort the typeface as soon as they become noticable. This is best demonstrated by using them as an exclusive adjustment mechanism:

Adjusting by letter spacing only (.letadj 96 100 12 110 100)

Harmony, liberal intercourse with all nations, are recommended by policy, humanity, and interest. But even our commercial policy should hold an equal and impartial hand; neither seeking nor granting exclusive favors or preferences; consulting the natural course of things; diffusing and diversifying by gentle means the streams of com-

Adjusting by glyph reshaping only (.letadj 100 96 100 110)

Harmony, liberal intercourse with all nations, are recommended by policy, humanity, and interest. But even our commercial policy should hold an equal and impartial hand; neither seeking nor granting exclusive favors or preferences; consulting the natural course of things; diffusing and diversifying by gentle means the streams of com-

In combination with adjustment of interword spaces, and if applied with rather strict limits, microtypography can have positive effects, though. This is especially true when lines are short; the sample text used so far can be formatted acceptably only if shrinking of interword spaces, paragraph-at-once adjustment, and microtypography are all combined:

Paragraph-at-once adjustment, word spacing 75 % to 150 %, letter spacing 95 % to 105 %, no glyph reshaping (.letadj 95 100 18 105 100)

Harmony, liberal intercourse with all nations, are recommended by policy, humanity, and interest. But even our commercial policy should hold an equal and impartial hand; neither seeking nor granting exclusive favors or preferences; consulting the natural course of things; diffusing and diversifying by gentle means the streams of commerce, but forc-

Paragraph-at-once adjustment, word spacing 75 % to 150 %, letter spacing 95 % to 105 %, glyph reshaping 98 % to 102 % (.letadj 95 98 18 105 102)

Harmony, liberal intercourse with all nations, are recommended by policy, humanity, and interest. But even our commercial policy should hold an equal and impartial hand; neither seeking nor granting exclusive favors or preferences; consulting the natural course of things; diffusing and diversifying by gentle means the streams of commerce, but forcing noth-

For layouts with longer lines than in this example, best results are normally achieved with even smaller ranges for letter spacing and glyph reshaping.

Penalties for line breaks and hyphenation

It is possible to tell *troff* that a line break after a certain word (or hyphenated word part) is preferred or discouraged by writing “\j’*N*’”. A positive *N* makes a breakpoint less likely, a negative *N* makes it more likely.

By default, *troff* hyphenates words whenever necessary in paragraph-at-once mode to minimize the adjustment. To reduce the number of hyphenations, *hyphenation penalties* can be configured. Then whenever a breakpoint involves a hyphenation, it is treated as less optimal, and another breakpoint that does not require a hyphenated word may be preferred even though more adjustment may become necessary.

Additional penalties can be configured for breakpoints that involve two successive hyphens, and for hyphenating the last word of a paragraph.

The “.hypp” request takes the single hyphen penalty as its first argument, the penalty for consecutive hyphens as second, and the penalty for hyphenating the last word as third. Effective penalties are between 10 and 200.

Hyphenation penalties only make hyphens less likely but do not limit them forcibly. The “.hlm” request imposes a strict limit on the number of consecutive hyphens. It causes a certain breakpoint to be disabled completely and can thus result in a non-optimal adjustment. It is recommended to use it in combination with “.spreadwarn” to detect such problems.

Paragraph shapes

The standard requests for setting indent and line length cannot be used inside a paragraph in paragraph-at-once adjustment mode. It is possible to define the shape of an entire paragraph line-by-line with the “.pshape” request. It takes a list of indent and line length pairs as arguments; the first pair applies to the first line of the paragraph, the second pair to the second line, and so forth. If the paragraph has more lines than pairs are given, the last pair is used for them; it it has less lines, the excess pairs are discarded. A paragraph shape is applied to a single paragraph only; it overrides the standard indent, temporary indent, and line length settings, of which indent and line length become effective for the next paragraph again.

For example, it is possible to create a paragraph whose shape forms a circle. To create a holey shape like this, define an indent and line pair for each contiguous part and use traps to move the resulting lines in vertical direction to the desired position. This is best done in a diversion so that the whole structure is kept together; diversion traps are the mechanism of choice then. ¶ Admittedly, creating a circle with a hole is hardly a serious application of the ".pshape" request. Complicated shapes almost always require a carefully wording of content, so paragraph-at-once formatting is only a limited aid when creating them. But ".pshape" is also needed to flow text around an image, even if it has simply a rectangular shape. ¶ If you know the "\parshape" command from TeX, note that the indent is included in the line length in *troff,* so you have to add every first value to every second one for reusing such shapes.

Notes

For paragraph-at-once adjustment, *troff* uses a variation of the algorithm originally developed by Donald Knuth and Michael Plass for the TeX system[2].

The criteria for the quality of a line differ: There is no explicit stretchability setting, and the total shrinkability is used to determine whether a breakpoint is feasible, but not for computing its optimality. *troff* currently has a slight preference for tight lines. It might make sense to have this configurable, but it seems that the fact that a line may be stretched or shrinked by a high amount does not necessarily indicate the optimality of doing so.

troff does not generate "overfull boxes", i.e. unadjustable text extending beyond the margin, unless the width of a single word exceeds that of the line. Its warning mechanism can inform about unacceptable adjustments.

troff makes no use of *fitness classes* and does not prefer to group lines of similar non-optimal spacing. This is because doing so may lead to a more even appearance of the lines of a paragraph, but at the expense of a less even appearance in the context of the whole document. For example, consider the case of multiple consecutive lines with loose spacing: If these lines are viewn in isolation, their spacing looks harmonically. If they are viewn as part of a document, they look more bright than the rest. It is not clear how to solve this without optimizing the spacing globally for a document, which is not a realistic option.

Breakpoints that might occur at different lines are currently not evaluated separately for each such line. Future evaluation may indicate that doing so is necessary in practice.

troff implements "microtypography" similarly as described by Hàn Thế Thành for *TEX*[3]. In particular, it performs a function like "level 2 font expansion" (p. 70), i.e. it considers the possibility to shrink interletter spaces and character shapes when computing breakpoints. Both stretchability and shrinkability are taken into account for computing the optimality of a breakpoint, but only as far as the width of the possible line is concerned; the percentual adjustment limits do not influence optimality.

The sample text is an excerpt of George Washington's 1796 Farewell address. Its choice for this purpose is due to examples in James Felici's *Complete Manual of Typography*[1].

References

[1] J. Felici, *The Complete Manual of Typography,* Berkeley, CA, 2003, pp. 147–149.

[2] D. E. Knuth, M. F. Plass, "Breaking paragraphs into lines", *Software—Practice and Experience*, Vol. 11, Issue 12 (1981), pp. 1119–1184; also in D. E. Knuth, *Digital Typography*, Stanford, 1999 (CSLI lecture notes no. 78), pp. 67–155.

[3] Hàn Thế Thành, *Micro-typographic extensions to the TEX typesetting system*, Masaryk University Brno, 2000.