Network Working Group B. Lilly
Internet-Draft May 2005
Intended status: Informational
Expires: November 28, 2005

Media subtype registration for media type text/troff
draft-lilly-text-troff-04

Status of this Memo

Copyright Notice

Abstract

A text media subtype for tagging content consisting of juxtaposed text and formatting directives as used by the troff series of programs and for conveying information about the intended processing steps necessary to produce formatted output is described.

Table of Contents

1. Introduction

It is sometimes desirable to format text in a particular way for presentation. One approach is to provide formatting directives in juxtaposition to the text to be formatted. That approach permits reading the text in unformatted form (by ignoring the formatting directives), and it permits relatively simple repurposing of the text for different media by making suitable alterations to the formatting directives or the environment in which they operate. One particular series of related programs for formatting text in accordance with that model is often referred to generically as "troff", although that is also the name of a particular lineage of programs within that generic category for formatting text specifically for typesetter and typesetter-like devices. A related formatting program within the generic "troff" category, usually used for character-based output such as (formatted) plain text is known as "nroff". For the purpose of the media type defined here, the entire category will be referred to simply by the generic "troff" name. Troff as a distinct set of programs first appeared in the early 1970s [N1.CSTR54], based on the same formatting approach used by some earlier programs ("runoff" and "roff"). It has been used to produce documents in various formats, ranging in length from short memoranda to books (including tables, diagrams, and other non-textual content). It remains in wide use as of the date of this document; this document itself was prepared using the troff family of tools per [I1.RFC2223] and [I2.Lilly05].

The basic format (juxtaposed text and formatting directives) is extensible and has been used for related formatting of text and graphical document content. Formating is usually controlled by a set of macros; a macro package is a set of related formatting tools, written in troff format (although compressed binary representations have also been used) and using basic formatting directives to extend and manage formatting capabilities for document authors. There are a number of preprocessors which transform a textual description of some content into the juxtaposed text and formatting directives necessary to produce some desired output. Preprocessors exist for formatting of tables of text and non-textual material, mathematical equations, chemical formulae, general line drawings, graphical representation of data (in plotted coordinate graphs, bar charts, etc.), representations of data formats, and representations of the abstract mathematical construct known as a graph (consisting of nodes and edges). Many such preprocessors use the same general type of input format as the formatters, and such input is explicitly within the scope of the media type described in this document.

2. Requirement Levels

The key words "MUST", "SHOULD", "SHOULD NOT", "RECOMMENDED", and "MAY" in this document are to be interpreted as described in [N2.BCP14].

3. Scope of Specification

The described media type refers to input which may be processed by preprocessors and by a page formatter. It is intended to be used where content has some text which may be comprehensible (either as text per se or as a readable description of non-text content) without machine processing of the content. Where there is little or no comprehensible text content, this media type SHOULD NOT be used. For example, while output of the "pic" preprocessor certainly consists of troff-compatible sequences of formatting directives, the sheer number of individual directives interspersed with any text that might be present makes comprehension difficult, whereas the preprocessor input language (as described in the Published Specification section of the registration below) may provide a concise and comprehensible description of graphical content. Preprocessor output which includes a large proportion of formatting directives would best be labeled as a subtype of the application media type. If particular preprocessor input content describes only graphical content with little or no text, and which is not readily comprehensible from a textual description of the graphical elements, a subtype of the image media type would be appropriate. The purpose of labeling media content is to provide information about that content to facilitate use of the content. Use of a particular label requires some common sense and judgment, and SHOULD NOT be mechanically applied to content in the absence of such judgment.

4. Registration Form

The registration procedure and form are specified in [I3.BCP13].

Type name: text

Subtype name: troff

Required parameters: none

Optional parameters:

charset: Must be a charset registered for use with MIME text types [N3.RFC2046], except where transport protocols are explicitly exempted from that restriction. Specifies the charset of the media content. With traditional source content, this will be the default "US-ASCII" charset. Some recent versions of troff processing software can handle Unicode input charsets, however there may be interoperability issues if the input uses such a charset (see Interoperability considerations below).

process: Human-readable additional information for formatting, including environment variables, preprocessor arguments and order, formatter arguments, and postprocessors. The parameter value may need to be quoted or encoded as provided for by [N4.RFC2045] as amended by [N5.RFC2231] and [N6.Errata]. Generating implementations must not encode executable content and other implementations must not attempt any execution or other interpretation of the parameter value, as the parameter value may be prose text. Implementations SHOULD present the parameter (after reassembly of continuation parameters, etc.) as information related to the media type, particularly if the media content is not immediately available (e.g. as with message/external-body composite media [N3.RFC2046]).

resources: Lists any additional files or programs that are required for formatting (e.g. via .cf, .nx, .pi, .so, and/or .sy directives).

versions: Human-readable indication of any known specific versions of preprocessors, formatter, macro packages, postprocessors, etc. required to process the content.

Encoding considerations:

7bit is adequate for traditional troff provided line endings are canonicalized per [N3.RFC2046]. Transfer of this media type content via some transport mechanisms may require or benefit from encoding into a 7bit range via a suitable encoding method such as the ones described in [N4.RFC2045]. In particular, some lines in this media type might begin or end with whitespace, and that leading and/or trailing whitespace might be discarded or otherwise mangled if the media type is not encoded for transport.

8bit may be used with Unicode characters represented as a series of octets using the utf-8 charset [I4.RFC3629], where transport methods permit 8bit content and where content line length is suitable. Transport encoding considerations for robustness may also apply, and if a suitable 8bit encoding mechanism is standardized, it might be applicable for protection of media during transport.

binary may be necessary when raw Unicode is used or where line lengths exceed the allowable maximum for 7bit and 8bit content [N4.RFC2045], and may be used in environments (e.g. HTTP [I5.RFC2616]) where Unicode characters may be transferred via a non-MIME charset such as UTF-16 [I6.RFC2781].

Restrictions on usage: none

Security considerations: Some troff directives (.sy and .pi) can cause arbitrary external programs to be run. Several troff directives (.so, .nx, and .cf) may read external files (and/or devices on systems which support device input via file system semantics) during processing. Several preprocessors have similar features. Some implementations have a "safe" mode which disables some of these features. Formatters and preprocessors are programmable, and it is possible to provide input which specifies an infinite loop which could result in denial of service, even in implementations which restrict use of directives that access external resources. Users of this media type SHOULD be vigilant of the potential for damage which may be caused by careless processing of media obtained from untrusted sources.

Processing of this media type other than by facilities which strip or ignore potentially dangerous directives, and processing by preprocessors and/or postprocessors SHOULD NOT be invoked automatically (i.e. w/o user confirmation). In some cases, as this is a text media type (i.e. it contains text which is comprehensible without processing), it may be sufficient to present the media type with no processing at all. However, like any other text media, this media type may contain control characters, and implementers SHOULD take precautions against untoward consequences of sending raw control characters to display devices.

Users of this media type SHOULD carefully scrutinize suggested command lines associated with the "process" parameter, contained in comments within the media, or conveyed via external mechanisms, both for attempts at social engineering and for the effects of ill-considered values of the parameter. While some implementations may have "safe" modes, those using this media type MUST NOT presume that they are available or active.

Comments may be included in troff source; comments are not formatted for output. However, they are of course readable in the troff document source. Authors should be careful about information placed in comments, as such information may result in a leak of information, or may have other undesirable consequences.

While it is possible to overlay text with graphics or otherwise produce formatting instructions that would visually obscure text when formatted, such measures do not prevent extracting text from the document source, and might be ineffective in obscuring text when formatted electronically, e.g. as PostScript or PDF.

Interoperability considerations: Recent implementations of formatters, macro packages, and preprocessors may include some extended capabilities which are not present in earlier implementations. Use of such extensions obviously limits the ability to produce consistent formatted output at sites with implementations that do not support those extensions. Use of any such extensions in a particular document using this media type SHOULD be indicated via the "versions" parameter value.

As mentioned in the Introduction, macro packages are troff documents, and their content may be subject to copyright. That has led to multiple independent implementations of macro packages, which may exhibit gross or subtle differences with some content.

Some preprocessors or postprocessors might be unavailable at some sites. Where some implementation is available, there may be differences in implementation that affect the output produced. For example, some versions of the "pic" preprocessor provide the capability to fill a bounded graphical object; others lack that capability. Of those that support that feature, there are differences in whether a solid fill is represented by a value of 0.0 vs. 1.0. Some implementations support only gray-scale output; others support color.

Preprocessors or postprocessors may depend on additional programs such as awk, and implementation differences (including bugs) may lead to different results on different systems (or even on the same system with a different environment).

There is a wide variation in the capabilities of various presentation media and the devices used to prepare content for presentation. Indeed, that is one reason that there are two basic formatter program types (nroff for output where limited formatting control is available, and troff where a greater range of control is possible). Clearly, a document which is designed to use complex or sophisticated formatting might not be representable in simpler media or with devices lacking certain capabilities. Often it is possible to produce a somewhat inferior approximation; colors might be represented as gray-scale values, accented characters might be produced by overstriking, italics might be represented by underlining, etc.

Various systems store text with different line ending codings. For the purpose of transferring this media type between systems or between applications using MIME methods, line endings MUST use the canonical CRLF line ending per [N3.RFC2046].

Published specification: [N1.CSTR54]

Applications which use this media type: The following applications in each sub-category are examples. The lists are not intended to be exhaustive.

Preprocessors: tbl [I7.CSTR49], grap [I8.CSTR114], pic [I9.CSTR116], chem [I10.CSTR122], eqn [I11.eqn], dformat [I12.CSTR142]

Formatters: troff, nroff, Eroff, sqtroff, groff, awf, cawf

Format converters: deroff, troffcvt, unroff, troff2html, mm2html

Macro packages: man [I13.UNIXman1], me [I14.me], mm [I15.DWBguide], ms [I16.ms], mv [I15.DWBguide], rfc [I2.Lilly05]

Additional information:

Magic number(s): None, however the content format is distinctive (see Published specification).

File extension(s): Files do not require any specific "extension". Many are in use as a convenience for mechanized processing of files, some associated with specific macro packages or preprocessors; others are ad-hoc. File names are orthogonal to the nature of the content. In particular, while a file name or a component of a name may be useful in some types of automated processing of files, the name or component might not be capable of indicating subtleties such as proportion of textual (as opposed to image or formatting directive) content. This media type SHOULD NOT be assigned a relationship with any file "extension" where content may be untrusted unless there is provision for human judgment which may be used to override that relationship for individual files. Where appropriate, a filename MAY be suggested by a suitable mechanism such as the one specified in [I17.RFC2183] as amended by [N5.RFC2231] and [N6.Errata].

Macintosh File Type Code(s): unknown

Person & email address to contact for further information:
Bruce Lilly
blilly@erols.com

Intended usage: COMMON

Author/Change controller: IESG

Consistency: The media has provision for comments; these are sometimes used to convey recommended processing commands, to indicate required resources, etc. To avoid confusing recipients, senders SHOULD ensure that information specified in optional parameters is consistent with any related information which may be contained within the media content.

5. Acknowledgments

The author would like to acknowledge the helpful comments provided by members of the ietf-types mailing list.

6. Security Considerations

Security considerations are discussed in the media registration. Additional considerations may apply when the media subtype is used in some contexts (e.g. MIME [I18.RFC2049]).

7. Internationalization Considerations

The optional charset parameter may be used to indicate the charset of the media type content. In some cases, that content's charset might be carried through processing for display of text. In other cases, combinations of octets in particular sequences are used to represent glyphs which cannot be directly represented in the content charset. In either of those categories, the language(s) of the text might not be evident from the character content, and it is RECOMMENDED that a suitable mechanism (e.g. [I19.RFC3282]) be used to convey text language where such a mechanism is available [I20.BCP18]. Where multiple languages are used within a single document, it may be necessary or desirable to indicate the languages to readers directly via explicit indication of language in the content. In still other cases, the media type content (while readable and comprehensible in text form) represents symbolic or graphical information such as mathematical equations or chemical formulae, which are largely global and language-independent.

8. IANA Considerations

IANA shall enter and maintain the registration information in the media type registry as directed by the IESG.

Appendix A. Examples

A.1. Data Format

The input:

```
Content-Type: text/troff ; process="dformat | pic -n | troff -ms"

Here's what an IP packet header looks like:
.begin dformat
style fill off
style bitwid 0.20
style recspread 0
style recht 0.33333
noname
 0-3 \0Version
 4-7 IHL
 8-15 \0Type of Service
 16-31 Total Length
noname
 0-15 Identification
 16-18 \0Flags
 19-31 Fragment Offset
noname
 0-7 Time to Live
 8-15 Protocol
 16-31 Header Checksum
noname
 0-31 Source Address
noname
 0-31 Destination Address
noname
 0-23 Options
 24-31 Padding
.end
```

produces as output:

Here's what an IP packet header looks like:

<table>
<tr><td colspan="4">Version
0 3</td><td colspan="4">IHL
4 7</td><td colspan="8">Type of Service
8 15</td><td colspan="16">Total Length
16 31</td></tr>
<tr><td colspan="16">Identification
0 15</td><td colspan="3">Flags
16 18</td><td colspan="13">Fragment Offset
19 31</td></tr>
<tr><td colspan="8">Time to Live
0 7</td><td colspan="8">Protocol
8 15</td><td colspan="16">Header Checksum
16 31</td></tr>
<tr><td colspan="32">Source Address
0 31</td></tr>
<tr><td colspan="32">Destination Address
0 31</td></tr>
<tr><td colspan="24">Options
0 23</td><td colspan="8">Padding
24 31</td></tr>
</table>

A.2. Simple Diagram

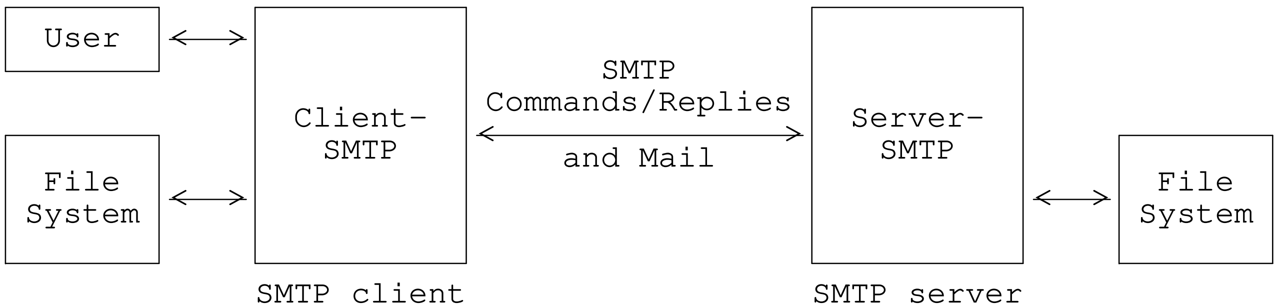
The input:

```
Content-Type: text/troff ; process="pic -n | troff -ms"

The SMTP design can be pictured as:
.DS B
.PS
boxwid = 0.8
# arrow approximation that looks acceptable in troff and nroff
define myarrow X A: [ move right 0.055;\
 "<" ljust;line right ($1 - 0.1);">" rjust;\
 move right 0.045 ]\
X
User: box ht 0.333333 "User"
FS: box ht 0.666667 "File" "System" with .n at User.s -0, 0.333333
Client: box ht 1.333333 wid 1.1 "Client\-" "SMTP" \
 with .sw at FS.se +0.5, 0
"SMTP client" rjust at Client.se -0, 0.166667
move to User.e ; myarrow(0.5)
move to FS.e ; myarrow(0.5)
move to Client.e ; SMTP: myarrow(1.8)
Server: box ht 1.333333 wid 1.1 "Server\-" "SMTP" \
 with .sw at Here.x, Client.s.y
box invis ht 0.5 "SMTP" "Commands/Replies" with .s at SMTP.c
box invis ht 0.25 "and Mail" with .n at SMTP.c
"SMTP server" ljust at Server.sw -0, 0.166667
move to Server.e.x, FS.e.y ; myarrow(0.5)
FS2: box ht 0.666667 "File" "System" \
 with .sw at Server.se.x +0.5, FS.s.y
.PE
.DE
```

produces as output:

The SMTP design can be pictured as:



Appendix B. Disclaimers

This document has exactly one (1) author.

In spite of the fact that the author’s given name may also be the surname of other individuals, and the fact that the author’s surname may also be a given name for some females, the author is, and has always been, male.

The presence of "or she", "/SHE", "each", "their", and "authors" (plural) in the boilerplate sections of this document is irrelevant. The author of this document is not responsible for the boilerplate text.

Comments regarding the silliness, lack of accuracy, and lack of precision of the boilerplate text should be directed to the IESG, not to the author.

Appendix C. Change History

[[This change history will not be part of a published RFC]]

-03 to -04 (IESG review and RFC preparation)

- Clarified application lists as examples, not exhaustive
- Added informative reference to RFC 2048 (BCP 13) (N.B. successor in process)
- Added clarification of extra security considerations
- Conformed draft reference to style suggested in I-D Checklist
- Revised text regarding command-line security issues

-02 to -03 (Last Call comments and final tweaks)

- revised description of process parameter
- cleaned up presentation of examples

- noted that RFC 2183 is amended by RFC 2231 and errata
- updated reference to using-troff draft
- added disclaimer of boilerplate content silliness, etc.
- added discussion of graphic overlays to security considerations

-01 to -02

- added this change history
- improved pic diagram
- revised boilerplate

-00 to -01

- revised reference numbering style
- informative reference to using-troff draft
- revised layout
- discussion of presentation w/o processing, also control character issues
- added emphasis to security considerations
- added information regarding comments vs. privacy/security
- improved data format layout
- added pic diagram (SMTP)

Normative References

[N1.CSTR54] Ossanna, Joseph F., "NROFF/TROFF User's Manual", Computing Science Technical Report No.54, Bell Laboratories, Murray Hill, New Jersey, 1976.

[N2.BCP14] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[N3.RFC2046] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types", RFC 2046, November 1996.

[N4.RFC2045] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies", RFC 2045, November 1996.

[N5.RFC2231] Freed, N. and K. Moore, "MIME Parameter Value and Encoded Word Extensions: Character Sets, Languages, and Continuations", RFC 2231, November 1997.

[N6.Errata] RFC-Editor errata page, http://www.rfc-editor.org/errata.html

Informative References

[I1.RFC2223] Postel, J. and J. Reynolds, "Instructions to RFC Authors", RFC 2223, October 1997.

[I2.Lilly05] Lilly, B., "Writing Internet-Drafts and Requests For Comments using troff and nroff" (draft-lilly-using-troff-04.txt), (draft-lilly-using-troff-04.ps), (draft-lilly-using-troff-04.pdf), May 2005.

[I3.BCP13] Freed, N., Klensin, J., and J. Postel, "Multipurpose Internet Mail Extensions (MIME) Part Four: Registration Procedures", BCP 13, RFC 2048, November 1996.

[I4.RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, RFC 3629, November 2003.

[I5.RFC2616] Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1", RFC 2616, June 1999.

[I6.RFC2781] Hoffman, P. and F. Yergeau, "UTF-16, an encoding of ISO 10646", RFC 2781, February 2000.

[I7.CSTR49] Lesk, M. E., "TBL - A Program for Setting Tables", Bell Laboratories Computing Science Technical Report #49, Murray Hill, New Jersey, 1976.

[I8.CSTR114] Bentley, Jon L. and Kernighan, Brian W., "Grap - A Language for Typesetting Graphs Tutorial and User Manual", Computing Science Technical Report No.114, AT&T Bell Laboratories, Murray Hill, New Jersey, 1991.

[I9.CSTR116] Kernighan, Brian W., "Pic - A Graphics Language for Typesetting User Manual", Computing Science Technical Report No.116, AT&T Bell Laboratories, Murray Hill, New Jersey, 1991.

[I10.CSTR122] Bentley, Jon L., Jelinski, Lynn W., and Kernighan, Brian W., "Chem - A Program for Typesetting Chemical Diagrams: User Manual", Computing Science Technical Report No.122, AT&T Bell Laboratories, Murray Hill, New Jersey, 1992.

[I11.eqn] Kernighan, Brian W, and Cherry, Lorinda L., "A System for Typesetting Mathematics", Communications of the ACM 18, 182-193, 1975.

[I12.CSTR142] Bentley, Jon L. "DFORMAT - A Program for Typesetting Data Formats", Computing Science Technical Report No.142, AT&T Bell Laboratories, Murray Hill, New Jersey, 1988.

[I13.UNIXman1] AT&T Bell Laboratories, "UNIX TIME-SHARING SYSTEM (VOLUME 1) : UNIX Programmer's Manual", Holt, Rinehart, & Winston, 1979

[I14.me] Allman, Eric P., "Writing Papers With NROFF Using -me", USD:19, University of California, Berkeley, Berkeley, California, 1997.

[I15.DWBguide] AT&T Bell Laboratories, "Unix System V Documenter's Workbench User's Guide", Prentice Hall, 1989

[I16.ms] Lesk, M. E., "Typing Documents on the UNIX System: Using the -ms Macros with Troff and Nroff", 1978, in "UNIX TIME-SHARING SYSTEM (VOLUME 2) : UNIX Programmer's Manual", Holt, Rinehart, & Winston, 1979

[I17.RFC2183] Troost, R., Dorner, S., and K. Moore, "Communicating Presentation Information in Internet Messages: The Content-Disposition Header Field", RFC 2183, August 1997.

[I18.RFC2049] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part Five: Conformance Criteria and Examples", RFC 2049, November 1996.

[I19.RFC3282] Alvestrand, H., "Content Language Headers", RFC 3282, May 2002.

[I20.BCP18] Alvestrand, H., "IETF Policy on Character Sets and Languages", BCP 18, RFC 2277, January 1998.

Author's Address

Bruce Lilly

Email: blilly@erols.com

Full Copyright Statement

Intellectual Property Statement

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.