Network Working Group
Internet Draft
Expires: June 5, 2005

B. Lilly
December 2004

Media subtype registration for media type text/troff
draft-lilly-text-troff-00

Status of this Memo

This document is an Internet-Draft and is subject to all provisions of section 3 of RFC 3667. By submitting this Internet-Draft, the author represents that any applicable patent or other IPR claims of which he is aware have been or will be disclosed, and any of which he become aware will be disclosed, in accordance with RFC 3668.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/1id-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

Copyright Notice

Copyright © The Internet Society (2004).

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.

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 [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 [RFC2223].

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 [RFC2119].

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 is certainly 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 Template

Type name:
text

Subtype name:
troff

Required parameters:
none

Optional parameters:

charset
Must be a charset registered for use with MIME text types [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
Lists a recommended command pipeline for formatting. The parameter value may need to be quoted or encoded as provided for by [RFC2045] as amended by [RFC2231] and [errata].

resources
Lists any required 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 [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 [RFC2045].

8bit may be used with Unicode characters represented as a series of octets using the utf-8 charset [RFC3629], where transport methods permit 8bit content and where content line length is suitable.

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

A command line, such as may be suggested via the optional "process" parameter, is a powerful tool when used by a computer-literate person. Individuals lacking basic security knowledge and/or common sense should not be given unsupervised access to a command line. Users of this media type should carefully examine the suggested command pipeline and media content

before executing commands.

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 version 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 [RFC2046].

Published specification:
[CSTR54]

Applications which use this media type:

Preprocessors
tbl [CSTR49], grap [CSTR114], pic [CSTR116], chem [CSTR122], eqn [eqn], dformat [CSTR142]

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

Format converters
deroff, troffcvt, unroff, troff2html, mm2html

Macro packages
man [UNIXman1], me [me], mm [DWBguide], ms [ms], mv [DWBguide]

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 [RFC2183].

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 in some contexts (e.g. MIME [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. [RFC3066]) be used to convey text language where such a mechanism is available. 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. Example

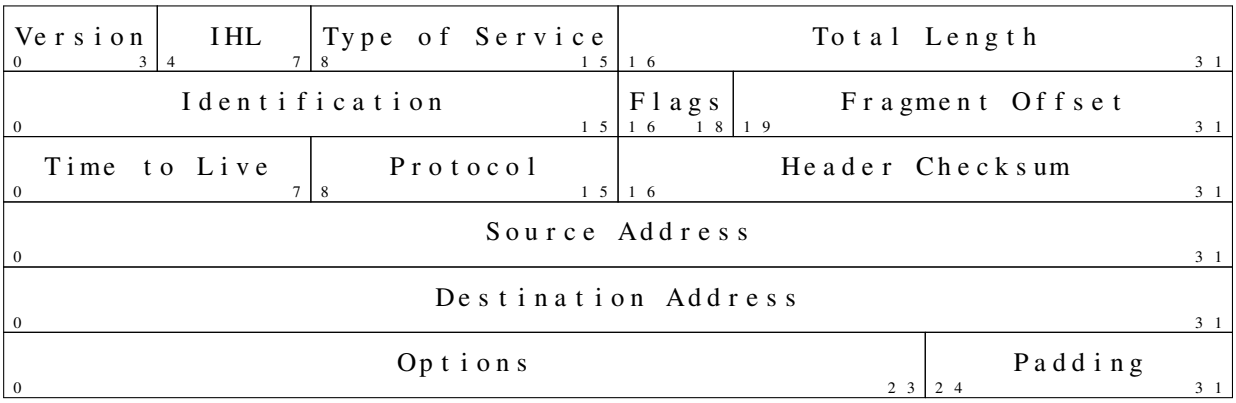
The input:

```
Content-Type: text/troff ;
 process="GROFF_NO_SGR=1 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 0.34
noname
 0-3 Version
 4-7 IHL
 8-15 Type of Service
 16-31 Total Length
noname
 0-15 Identification
 16-18 Flags
 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:



Normative References

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

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

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

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

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

Informative References

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

[CSTR97] Kernighan, Brian W., "A Typesetter-independent TROFF", Computing Science Technical Report No.97, AT&T Bell Laboratories, Murray Hill, New Jersey, 1982.

[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.

[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.

[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.

[CSTR132] Bentley, Jon L. and Kernighan, Brian W., "A System for Algorithm Animation Tutorial and User Manual", Computing Science Technical Report No.132, AT&T Bell Laboratories, Murray Hill, New Jersey, 1991.

[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.

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

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

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

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

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

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

[RFC2277] Alvestrand, H., "IETF Policy on Character Sets and Languages", BCP 18, RFC 2277, January 1998.

[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.

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

[RFC3066] Alvestrand, H., "Tags for the Identification of Languages", BCP 47, RFC 3066, January 2001.

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

[RFC3667] Bradner, S., "IETF Rights in Contributions", BCP 78, RFC 3667, February 2004.

[RFC3668] Bradner, S., "Intellectual Property Rights in IETF Technology", BCP 79, RFC 3668, February 2004.

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

Author's Address

Bruce Lilly

Email: blilly@erols.com

Full Copyright Statement

Copyright © The Internet Society (2004). This document is subject to the rights, licenses and restrictions contained in BCP 78 [RFC3667], and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights

might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 [RFC3667] and BCP 79 [RFC3668].

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Acknowledgment

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