

Simplemail: A Simple Standardized Markup for Electronic Mail and Usenet Messages

1 Status of this Memo

This RFC suggests a proposed protocol for the Internet community, and requests discussion and suggestions for improvements. The protocol is called “simplemail”, and is intended as a simple text markup system for mail messages.

Distribution of this memo is unlimited.

2 Introduction

2.1 Relation to MIME

This document is meant to supplement RFC 1341 “MIME (Multipurpose Internet Mail Extensions)”, which extends the notion of an electronic mail message to allow messages comprising multiple parts, each in various formats (such as text, graphics, audio, or programs). Each such *Content-Type* may have various *subtypes* which further describe the standard to which such parts may adhere.

Simplemail is a proposed subtype of “Content-Type: Text”, which is meant to be used for messages which are primarily textual.

Electronic Mail, or email, is by nature a written medium, but due to limitations of the current standards, email authors cannot currently take advantage of the sorts of textual markup that are common in other forms of written or printed discourse. Such techniques include underlining or changing font to signal emphasis, to introduce new terms, to denote the titles of books, or to set off code fragments. It also includes indented displays, bulleted lists, footnotes, and section headings.

2.2 Description

MIME initially contained a specification for an SMGL-like subtype of Text called *richtext*. Unfortunately, *richtext* as defined is a very intrusive notation. It would be fine if it were purely a transport mechanism: if composition agents converted to it and user agents understood it and formatted the message accordingly. Unfortunately, that will not be the case for some time, and we must take into account the fact that many people will be using current UAs in the foreseeable future.

In the interim, what seems to be needed is a non-intrusive notation, one which will not obscure the message for people using old UAs (and may even hint at the writer's intent) and which will be easy to insert for people composing messages on old UAs. (The reading and composition agents are often different, and we can expect widespread use of more sophisticated readers before more sophisticated composers.)

Fortunately, due to the large numbers of people faced with the current email limitations, several such markup conventions have arisen. In this document, we attempt to formalize and standardize some of them. We also invent some notations where no current standard exists or where the current conventions do not readily lend themselves to mechanization.

One way in which we differ from richtext is that we focus on marking the *semantic category* of the text rather than its *form*. Thus while richtext had commands for manipulating fonts (such as "<bold>" and "<smaller>"), simplemail assigns things to categories such as *emphasized*, *alternate font*, or *literal*. We feel that it is up to the receiver's UA to decide the details of presentation.

We also attempt to preserve some of the structure of the document and allow writers to provide some structure to their text. We formalize the notion of *quoting* a message. We mark sections of the message such as *header lines*, *attribution lines*, and *signature lines*, and we allow the writer to insert *section headings*, *literal lines*, and *indented paragraphs* (among other things).

2.3 Philosophy and Restrictions

In defining simplemail, we have kept several guidelines in mind:

- A simplemail message should be trivial to type on a UA that gives little or no help (such as a plain text editor).
- As much as possible, it should mimic the current email and news conventions.
- It should be unlikely that a user would accidentally imply markup.
- It should be possible to apply the simple formatting that people are used to in handwritten and typewritten letters and email, but not the fancy control provided by word processors and document formatters.
- It should be possible to read a simplemail message as straight text on a non-compliant UA and get a good feel for what the markup means.
- It should be simple to display a message on a UA that understands the format, with varying levels of markup. That is, proportional fonts and faces such as bold and italics may be used when available, but are not necessary. The end user should be able to decide how the markup looks.
- It must be simple to include a block of literal text such as a file or program fragment.
- It must be trivial to quote non-simplemail messages in a simplemail message (and vice versa). Since netiquette requires that quoted material be

trimmed as much as possible, markup must be localized so that the partial deletion of markup information (such as the loss of one of two delimiters) should not have a disastrous effect.

3 Overview

This scheme introduces six pieces of markup, some of which are slightly parameterized. Three (inline markup, quoting, and indentation) are very close to current conventions. One (notes) is similar to an infrequent current usage. The remaining two (paragraph prefixes and line prefixes) are inventions, but sit in the left margin where they are unobtrusive.

For inline markup, phrases are bracketed by `ASTERISKS` to signal emphasis, `UNDERSCORES` to signal words that should stand out (neologisms and titles), and `BACKQUOTES` to signal literal or computer text.

Quoting is signaled by `GREATER THAN` in the left margin, optionally prefixed by an alphanumeric identifier to disambiguate the author or source of the quote.

Indentation of paragraphs is signalled by indenting the lines of the paragraph four `SPACES` for each indent level.

A paragraph may be preceded by a *paragraph prefix* to number or bullet the paragraph. The paragraph prefix begins with an `EQUAL SIGN` and lasts until the following space. An `EQUAL SIGN` by itself represents a bullet.

Lines which should (or may usefully) be treated specially by the UA are prefixed by a `COLON`. Such lines are generally not folded in with the rest of the paragraph. A second character may identify a particular type of line, such as literal, section heading, signature, or attribution.

Footnotes, citations, or external references are delimited by pairs of `ANGLE BRACKETS`. The note mark and its body may also be separated by placing a special tag at the beginning of the body paragraph.

The following sections explain this markup in more detail.

4 Markup

4.1 Text Lines and Paragraphs

Text in a simplemail message is formatted into filled paragraphs by default. Line breaks in the text are generally treated as `SPACE` characters, except in special cases mentioned in sections 4.3, 4.5, and 4.6. Whether text is actually filled, as well as the initial left and right margins, the first line indent, and whether the text is justified or hyphenated, is left up to the mail user agent. Actual simplemail source lines should contain fewer than 78 characters.

A paragraph ends with a blank line or a change of quotation prefix or indent level (see sections 4.3 and 4.6). A blank line is a line that contains only whitespace characters. Multiple paragraph ends in a row are merged into a single paragraph end.

4.2 Inline markup

Inline markup consists of a short sequence of words beginning and ending with some markup character, as in the following

```
this text is *marked for* emphasis so that read-
ers...
```

which contains markup for both *emphasis* and *alternate vocabulary*.

The opening delimiter must not be followed by whitespace, and the closing delimiter must not be preceded by whitespace.

Inline markup does not extend across paragraph boundaries. The markup characters may be used in text by doubling them, i.e.

```
this illustrates the use of ** asterisks ** in text,
and *even ** in ** emphasized* text.
```

4.2.1 Emphasized text.

The markup character for emphasis is the ASTERISK. This markup should be used to indicate stress or emphasis on some word or phrase.

```
I think that simplemail is *terrific*!
```

4.2.2 Alternate vocabulary text.

The markup character for words or phrases in an alternate vocabulary is the UNDERSCORE character. This markup should be used to indicate that a word or phrase is to be considered in a different way than the surrounding language or words, e.g., words in a different language, names of ships or books, technical terms.

```
We call this a delimiter.
```

4.2.3 Literal text.

The delimiter for literal text is BACKQUOTE. Within literal text no other markup is noticed. This is intended to be used for short code fragments or mentions of punctuation marks. When possible, literal text will probably be displayed in a contrasting font.

Literal text may span line breaks; the newlines are treated as spaces. To insert multi-line literals where line breaks are important, use *literal lines* described in section 4.5.1. After a newline within a literal, anything that looks like a quote prefix (see section 4.6) is stripped and compared to see if the paragraph has ended. An unterminated literal ends at the end of the paragraph (blank line or change of quote prefix).

```
If `a < b*2`, we call this...
```

4.3 Indentation

Paragraphs in Simplemail may be indented to allow displays, excerpts from books,

or (possibly nested) lists. The actual amount of indentation displayed is up to the reader's UA; the composer merely specifies the *level* of indentation for the paragraph.

A paragraph is indented one level for each four SPACES at the beginning of each of its lines¹ (following the quote prefix). Consecutive lines at different indentation levels constitute a paragraph break. TABS are explicitly deprecated, but if seen at the beginning of the line should be treated as equivalent to four SPACES (one indent level).

```
This is a paragraph at indentation level zero (the
main text).
....This is a paragraph at indentation level one. It
....may span multiple lines.
.....Now we're at level two.
This paragraph is back at level zero.
```

4.4 Paragraph Prefixes

If the first character of a paragraph begins with an EQUAL SIGN, the characters following the EQUAL SIGN up to the following SPACE are taken as a *prefix* for the paragraph. If the EQUAL SIGN is immediately followed by a SPACE, the prefix is taken to be a bullet character (or the UA's best approximation).

This allows the writer to easily specify numbered or bulleted lists.

```
=1. This is an enumerated list.

....=a) As is this, but the prefixes are alphabetic
....and we're indented another level.

=2. This is the outer list.

....= This is a bulleted list.

....= Another bulleted item.
```

Note that simplemail does not have any actual notion of an "enumerated list" per se, but rather paragraphs with numbered prefixes. There is no provision for automatically generating numbers. One of the primary reasons for this was the fear that messages quoting only part of a list would cause quoted paragraphs to change their numbers or lose entirely the fact that they were numbered, which could become very confusing.

4.5 Special Lines

Simplemail assumes that most line breaks (except those that signal the ends of paragraphs) are merely whitespace, and a UA is allowed to refold paragraphs. Occasionally, line breaks can be very important. Simplemail provides a way to say

1. That is, zero through three spaces is indent level zero, four through seven spaces is indent level one, etc. This implies that if exactly four spaces are used per level, the first line of a paragraph may be indented up to three extra spaces. Such indentation will probably be ignored by the reader's UA.

“this line should be treated as a unit and not folded in with the text that precedes or follows it”.

All unfilled lines begin with a COLON. In the default case, the colon is followed by one or more spaces. In the sections below we will see several semantically meaningful types of unfilled line whose prefixes are COLON *some character* SPACE.

```
Mark Twain's works include:
.....: _Huckleberry Finn_
.....: "The Mysterious Stranger"
.....: _Tom Sawyer_
.....: ...
```

4.5.1 Literal Lines

Often it is necessary to include a block of text in which not only is the line breaking significant, but the text may also not adhere to simplemail formatting conventions and/or may contain layout that assumes that it is being displayed in a fixed-width font (as with character graphics or tables). Common examples of this include code fragments, files,² or quotations of non-simplemail messages.

A literal quotation begins with COLON SINGLE-QUOTE SPACE. Unlike the other prefixes, whitespace following the prefix (except for the single SPACE) is retained. This allows code to be included while preserving each line's indentation. If at all possible, the UA should display literal lines in a fixed-width font to preserve the intended appearance.

```
What's wrong with this code?
.....:' int a[10];
.....:' for (i=0; i<=10; i++)
.....:' {
.....:' ..a[i] = i;
.....:' }
```

4.5.2 Section Titles

Many articles are naturally divided into *sections*. In simplemail, a section title is a paragraph that begins with COLON COLON SPACE, optionally followed by a PERIOD-separated list of digit strings, e.g. “5.2.2.8”, which indicate the depth of the section. Compliant UAs can choose an appropriate style for sections and can even extract a “Table of Contents”. Users of non-compliant UAs benefit as well, as the standardized prefix makes it possible to search for a section in most UAs or editors.

```
:: Introduction
:: 1.4 Where to go for help
```

4.5.3 Envelope Lines

Simplemail provides the ability to flag certain lines as being part of the *envelope* of

2. Although files are more properly be sent as MIME attachments.

the message (or a quoted message) as opposed to the body. Such lines include *signature lines*, *attribution lines*, and *header lines*.³ Normally these lines are inserted automatically by the composition agent. They are flagged primarily so that the receiving UA may display them specially, ideally deemphasizing them relative to the actual content of the message.

Signature lines are prefixed by COLON S SPACE, attribution lines by COLON A SPACE, and header lines by COLON H SPACE.

```
:A In article <1234567@berkeley.edu> clinton@whiteh...
> :H Message-Id: <1234567@berkeley.edu>
> :H From: clinton@whitehouse.gov
>
> Hi there. We're finally on the net. I look forward to
> talking to you all
>
> :S William Jefferson Clinton
> :S The Prez
> :S
> :S Don't blame me, I voted for Bill and Opus.
```

Looking at the Message-Id, I'd say that this is a pretty pathetic attempt at a hoax, even for a weenie.

```
:S Joe Physics
:S Stanford University
:S physics-god@stanford.edu
```

4.6 Quoting

One of the most common forms of markup in email today is quoting other articles. The quoted article may in turn quote other articles, and the nesting can get quite deep. Currently, the standard convention is to insert a *prefix* (generally GREATER-THAN) on each quoted line. Deeper quotes will have more than one prefix (as, “>>”).

In simplemail, each line of a quotation should begin with a *quotation prefix*, consisting of an optional string of alphanumeric characters, followed by the characters GREATER-THAN SPACE. The leading string may be used to differentiate multiple quotations in the same message, but may also be omitted.

```
This text is from the message being written
```

```
> This text is from a quoted message
```

```
Jim> This text is quoted from different message.
```

Often quotations are themselves quoted, and simplemail supports this as well, as arbitrarily many quotation prefixes may be present. The SPACE following the GREATER-THAN symbol may be omitted on all but the final prefix.

3. For quoted messages. The header lines of a simplemail message itself are, of course, marked by their position in the message

```

:A In article <9876@stanford.edu> physics-god@stanf...
> :A In article <1234567@berkeley.edu> clinton@whit...
> > :H Message-Id: <1234567@berkeley.edu>
> > :H From: clinton@whitehouse.gov
>
> Looking at the Message-Id, I'd say that this is a
> pretty pathetic attempt at a hoax, even for a weenie.

```

You expected more?

The quotation prefix is stripped before looking for blank lines, indentation prefixes, or unfilled-line symbols. A change in quotation prefix from one line to the next signifies the end of a paragraph.

See section 6.5 for hints about how to display quoted material.

When composing a reply or follow-up to a message formatted with simplemail, the editor should normally insert GREATER-THAN SPACE before each line. If the message was not in simplemail format, the editor should normally insert GREATER-THAN SPACE COLON QUOTE SPACE (“> :’ ”) to signal to UAs that they should not try to further interpret the line, fill paragraphs, etc. Header lines, if included should be flagged COLON H SPACE.

4.7 Handling References

4.7.1 Footnotes

Simplemail provides a footnote capability which provides for both in-line and out-of-line footnotes. These can be used both for commentary and pointers to reference materials.

An in-line footnote has the form LESS-THAN LESS-THAN *text* GREATER-THAN GREATER-THAN, as in

```
The main thesis<<Which I'd hate to have to defend>> is
```

A UA should extract such notes and display them elsewhere (below the current paragraph, at the end of the message, at the bottom of the page, via a hypertext link) and leave some sort of mark.

The writer may also provide his own number as a hint to the UA by prefixing the text with *digits* COLON SPACE, as in

```
This idea<<12: Due to Swanson>> is one of the best that
```

An out-of-line footnote is composed of two parts, the *mark* and the *note body*. The mark is of the form LESS-THAN LESS-THAN *digits* GREATER-THAN GREATER-THAN, and should be included in the text.

```
The main authority in the field<<1>> says that
```

A UA may want to replace this with a superscript or hypertext link.

The note body is a paragraph which begins with LESS-THAN LESS-THAN *digits* COLON GREATER-THAN GREATER-THAN. It may occur anywhere and should be extracted and displayed as are the bodies of in-line footnotes.

```
<<1:>> Don Knuth,  
: _The Art of Computer Programming_, vol. 1,  
: Addison Wesley, 1973
```

Quoted material may have its own notes. The UA should keep this straight.

Note that it is the responsibility of the writer (or the writer's composition agent) to number out-of-line notes.

4.7.2 External References

Simplemail allows references to messages, message-parts, and other documents outside of the message. The form LESS-THAN LESS-THAN ASTERISK SPACE *external-body-reference* GREATER-THAN GREATER-THAN should be used, where *external-body-reference* is a list of parameters as specified for the MIME type "message/external-body" in section 7.3.3 of RFC 1341. E.g.,

```
<<* access-type="anon-ftp",site="ftp.nisc.sri.com",  
directory="rfc",name="rfc1352.txt">>
```

The ASTERISK may be prefixed by a number to provide a fixed mark.

5 Rules for composing

There are only a few rules that need to be followed to ensure that a message follows this standard. Most of them can be automatically built into the composition agent (even if it is just implemented as a set of emacs functions):

1. When quoting a message, use a valid quote prefix. This must be a string of alphanumeric characters (at the left margin) followed by GREATER-THAN SPACE. GREATER-THAN SPACE alone is perfectly acceptable and encouraged.
2. If an attribution line is generated (as it should), prefix the line with COLON A SPACE. The content of the line is otherwise immaterial.
3. If headers are quoted, prefix each header line with COLON H SPACE.
4. When quoting non-simplemail messages, prefix each line of the body of the message with COLON QUOTE SPACE to flag it as a literal line.
5. When a signature is inserted in the message, ensure that each line starts with COLON S SPACE.

6 Hints for display

Simplemail is designed such that a UA can take advantage of whatever facilities are available to nicely display a message, but nothing is assumed. *more(1)* by itself is a compliant, though limited, simplemail viewer.

A more sophisticated viewer might run the message through a simple preprocessor, formatting the result for display in a fixed-width-font terminal emulator.

Still more sophisticated viewers might take advantage of the ability to change fonts and use variable-width fonts.

The most sophisticated viewers might be able to set footnotes at the bottom of the page or as the targets of hypertext buttons, present tables of contents for sectioned messages, or present the attribution line for selected text.

This section contains some of our hints for displaying the various markup features on different levels of viewer. With few exceptions, the markup characters themselves should be removed.

6.1 Paragraphs

Lines in paragraphs should be folded. If possible, a variable-width font should be used. The length of the line should be under user control. The first line of a paragraph should not receive any extra indentation, and space should be inserted between paragraphs (except perhaps when the paragraph ends due to changes in quotation prefix).

6.2 Inline Markup

Emphasized text should be displayed in italics or underlined (or in some other contrasting style such as bold or inverse). If no such features are available, the asterisks should be left on.

Alternate vocabulary text should be displayed in italics or underlined.⁴ If no such features are available, the underscores should be left on. .

Literal text should be displayed in a contrasting font, preferably a fixed-width or typewriter-like font. If no such features are available, the backquotes should be left on (or changed to some other form of quotation mark). Alternatively, the user should be able to choose to have literals displayed as normal text.

6.3 Indentation

Indented paragraphs should have their left margin increased to match the indent level. The UA may also want to narrow the right margin of an indented paragraph and perhaps alter the font size or line spacing.

Paragraph prefixes should be displayed as hanging tags in the margin to the left of the first line of the paragraph.

6.4 Unfilled Lines

All unfilled lines should not be folded with the lines above or below them in their paragraph. Except for literal lines, white space following the prefix is unnecessary, and should be stripped.

Literal lines should be displayed in a fixed-width font. Initial white space should be retained. Tabs should be expanded to multiples of eight character positions and converted to spaces *before* the paragraph's quote prefix and indentation are taken

4. The ambiguity between these two cases does not usually cause problems in other print media, in which they are routinely rendered the same way

into account. It is up to the UA what to do with long literal lines.

Section titles should be displayed in a larger, bold font. If this is not available, they should be set off by a line of dashes underlining the text. The UA should provide a way to search for the next, the previous, or a specific title, and may want to create a table of contents.

Envelope information such as header lines, attribution lines, and signature lines should in general be de-emphasized by rendering it in a smaller font. The UA may want to take advantage of the keyword/value nature of header lines to display the different parts (or different headers) in different styles (lining up all of the values, or emphasizing subjects and authors). It would be nice if attribution lines could be displayed in such a way that it was more easily apparent what text was being attributed.

6.5 Quoting

The quotation prefix (or some normalized form of it) should be displayed either in the left margin of each line, or at the beginning of the quotation, as an aid in following the conversation.

The UA may also want to display quotations from different articles in contrasting styles. A useful approximation is to use different colors for odd-numbered and even-numbered levels.

The quote prefix should be prepended to white space implied by the indent level.

6.6 Footnotes

Footnote marks should be displayed as superscripts or hypertext buttons in the text, or by some other means such as a bracketed number in a contrasting font. They may also be left as is.

Footnote bodies should be extracted and displayed elsewhere, probably either at the bottom of the page, at the end of a hypertext link, at the end of the document, or below the current paragraph. The UA must take into account that text at various quoting levels may have identically numbered footnotes and should act to disambiguate them.

The UA should assign a mark to in-line footnotes. If it chooses to use a number, it should ensure that it does not use one that the author uses. Even if some notes are missing at a particular quote level, the UA should not renumber notes.

It may be a good idea to number notes from the series “*”, “**”, “***”, etc. (or “*”, “†”, “‡”, etc. if available) to differentiate the two series. The only problem with this is that these characters are *not* allowable in out of line notes, and users may get confused.

6.7 Handling Missing Information.

Especially when messages get trimmed in quotation, important information may be lost. Due to the locality of information in simplemail, the UA should be able to

```

From: evan@hplerk.hpl.hp.com (Evan Kirshenbaum)
Organization: Hewlett-Packard Laboratories
Subject: Re: Simplemail Proposal

:A In article <456@parc.xerox.com> Bill Jassen writes:
> :A In article <789@hpl.hp.com> Evan Kirshenbaum writes:
> > So what do you think?
> I think we're getting very close.

Wow! If *we* can agree on this, convincing everyone else
should be a snap!

> We still have to
>   = Finish up the document.
Pretty close to done.
>   = Put it on an ftp site. Do you think we should do
>   that on my end our yours?
Probably easier done on your end.

> > What do you think about allowing `:2' for multi-column
> > stuff?
> >   : ' :2 word1   word2
> Let's see whether anybody wants it. We'd also have to
> agree on how to separate the columns.

Ok. I'll send you the latest draft.

:S Evan Kirshenbaum
:S kirshenbaum@hpl.hp.com

```

Figure 1. Simplemail Source

recover gracefully from this situation and often even recover the lost information. Even if some of the message is improperly displayed, in no case should a problem last past the end of the current paragraph.

Because opening in-line markup delimiters cannot be followed (and closing delimiters cannot be preceded) by white space, the UA can tell whether the first instance of a delimiter in a paragraph is supposed to be an open or a close. If it is a close, this probably meant that the paragraph was broken in the midst of the markup, and it can safely insert the opening delimiter at the beginning of the paragraph.

Similarly, if the paragraph ends before a closing delimiter is seen, the delimiter may justly be inserted at the end of the paragraph.

The UA must be prepared to deal with footnote marks whose bodies have been deleted as well as bodies whose marks have been deleted.

7 Summary of syntax

8 Extended example

This section contains a sample Usenet article and several possible displays. The original simplemail source is shown in figure 1. With a simple text filter, it might turn into something like figure 2. A more sophisticated UA might display it something like figure 3.

```
From:      evan@hplerk.hpl.hp.com (Evan Kirshenbaum)
Organization: Hewlett-Packard Laboratories
Subject:   Re: Simplemail Proposal

In article <456@parc.xerox.com> Bill Jassen writes:
> In article <789@hpl.hp.com> Evan Kirshenbaum writes:
> > So what do you think?
> > I think we're getting very close.

Wow! If *we* can agree on this, convincing everyone else
should be a snap!

> We still have to
>   o Finish up the document.

Pretty close to done.

>   o Put it on an ftp site. Do you think we should do
>     that on my end our yours?

Probably easier done on your end.

> > What do you think about allowing `:2` for multi-column
> > stuff?
> >   :2 word1      word2

> Let's see whether anybody wants it. We'd also have to
> agree on how to separate the columns.

Ok. I'll send you the latest draft.

Evan Kirshenbaum
kirshenbaum@hpl.hp.com
```

Figure 2. Filtered Display

9 Authors' Addresses

Bill Jassen
Xerox Corporation
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
Phone: (415) 812-4763
Email: jassen@parc.xerox.com

Evan Kirshenbaum
Hewlett-Packard Laboratories
3500 Deer Creek Road, 26U
Palo Alto, CA 94304
Phone: (415) 857-7572
Email: kirshenbaum@hpl.hp.com

From: evan@hplerk.hpl.hp.com (**Evan Kirshenbaum**)
Organization: Hewlett-Packard Laboratories
Subject: *Re: Simplemail Proposal*

In article <456@parc.xerox.com> Bill Jassen writes:
> In article <789@hpl.hp.com> Evan Kirshenbaum writes:
>> So what do you think?
> I think we're getting very close.
Wow! If *we* can agree on this, convincing everyone else should be a snap!
> We still have to
> • Finish up the document.
Pretty close to done.
> • Put it on an ftp site. Do you think we should do that on my end or yours?
Probably easier done on your end.
>> What do you think about allowing :2 for multi-column stuff?
>> :2 word1 word2
> Let's see whether anybody wants it. We'd also have to agree on how to separate the
> columns.
Ok. I'll send you the latest draft.
Evan Kirshenbaum
kirshenbaum@hpl.hp.com

Figure 3. Sophisticated Display