Internet Engineering Task Force (IETF)
Request for Comments: 5538
Category: Standards Track
ISSN: 2070-1721

F. Ellermann
xyzzy

April 2010

The 'news' and 'nntp' URI Schemes

Abstract

This memo specifies the 'news' and 'nntp' Uniform Resource Identifier (URI) schemes that were originally defined in RFC 1738. The purpose of this document is to allow RFC 1738 to be made obsolete while keeping the information about these schemes on the Standards Track.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/5538.

Copyright Notice

Table of Contents

1. Introduction

The first definition for many URI schemes appears in [RFC1738]. This memo extracts the 'news' and 'nntp' URI schemes from it to allow that material to remain on the Standards Track if [RFC1738] is moved to "historic" status. It belongs to a series of similar documents like [RFC4156], [RFC4157], [RFC4248], and [RFC4266], which are discussed on the <mailto:uri@w3.org> list.

The definitions for the 'news' and 'nntp' URI schemes given here are updates from [RFC1738] based on modern usage of these schemes. This memo intentionally limits its description of the 'news' URI scheme to essential features supposed to work with "any browser" and Network News Transfer Protocol (NNTP) server.

[RFC3986] specifies how to define schemes for URIs; it also explains the term "Uniform Resource Locator" (URL). The Network News Transfer Protocol (NNTP) is specified in [RFC3977]. The Netnews Article Format is defined in [RFC5536].

The key word "MUST" in this memo is to be interpreted as described in [RFC2119]. UTF-8 is specified in [RFC3629]. The syntax uses the ABNF defined in [RFC5234].

2. Background

The 'news' and 'nntp' URI schemes identify resources on an NNTP server, individual articles, individual newsgroups, or sets of newsgroups.

User agents like Web browsers supporting these schemes use the NNTP protocol to access the corresponding resources. The details of how they do this, e.g., employing a separate or integrated newsreader, depend on the implementation. The default <port> associated with NNTP in [RFC3977] is 119.

2.1. 'nntp' URIs

The 'nntp' URI scheme identifies articles in a newsgroup on a specific NNTP server. In [RFC3986] terminology, this means that 'nntp' URIs have a non-empty <authority> component; there is no default <host> as for the 'file' or 'news' URI schemes.

Netnews is typically distributed among several news servers, using the same newsgroup names but local article numbers. An article available as number 10 in group "example" on server "news.example.com" has most likely a different number on any other

server where the same article is still available. Users allowed to read and post articles on "their" server may not be allowed to access articles on an "arbitrary" server specified in an 'nntp' URI.

For these reasons, the use of the 'nntp' URI scheme is limited, and it is less widely supported by user agents than the similar 'news' URI scheme.

2.2. 'news' URIs

The 'news' URI scheme identifies articles by their worldwide unique "Message-ID", independent of the server and the newsgroup. Newsreaders support access to articles by their "Message-ID", without the overhead of a URI scheme. In simple cases, they do this directly as an NNTP client of a default or currently used server as configured by the user. More general user agents use the 'news' URI scheme to distinguish "Message-IDs" from similar constructs such as other URI schemes in contexts such as a plain text message body.

The 'news' URI scheme also allows the identification of newsgroups or sets of newsgroups independent of a specific server. For Netnews, a group "example" has the same name on any server carrying this group, exotic cases involving gateways notwithstanding. To distinguish "Message-IDs" and newsgroup names, the 'news' URI scheme relies on the "@" between local part (left-hand side) and domain part (right-hand side) of "Message-IDs".

[RFC1738] offered only one wildcard for sets of newsgroups in 'news' URIs, a "*" used to refer to "all available newsgroups". In common practice, this was extended to varying degrees by different user agents. An NNTP extension known as <wildmat>, specified in [RFC2980] and now part of the base NNTP specification, allows pattern matching in the style of the [POSIX] "find" command. For the purpose of this memo, this means that some additional special characters have to be allowed in 'news' URIs, some of them percent-encoded as required by the overall [RFC3986] URI syntax. User agents and NNTP servers not yet compliant with [RFC3977] do not implement all parts of this new feature.

Another commonly supported addition to the [RFC1738] syntax is the optional specification of a server at the beginning of 'news' URIs. This optional <authority> component follows the overall [RFC3986] syntax, preceded by a double slash "//" and terminated by the next slash "/", question mark "?", number sign "#", or the end of the URI.

2.3. Query Parts, Fragments, and Normalization

Fragments introduced by a number sign "#" are specified in [RFC3986]; the semantics is independent of the URI scheme, and the resolution depends on the media type.

This memo doesn't specify a query part introduced by a question mark "?" for the 'news' and 'nntp' URI schemes, but some implementations are known to use query parts instead of fragments internally to address parts of a composite media type [RFC2046] in Multipurpose Internet Mail Extensions (MIME).

There are no special "." or ".." path segments in 'news' and 'nntp' URLs. Please note that "." and ".." are not valid <newsgroup-name>s.

URI producers have to percent-encode some characters as specified below (Section 4); otherwise, they MUST treat a "Message-ID" without angle brackets for 'news' URLs as is, i.e., case-sensitive.

3. Syntax of 'nntp' URIs

An 'nntp' URI identifies an article by its number in a given newsgroup on a specified server, or it identifies the newsgroup without article number.

```
nntpURL        = "nntp:" server "/" group [ "/" article-number ]
server         = "//" authority              ; see RFC 3986
group          = 1*( group-char / pct-encoded )
article-number = 1*16DIGIT                   ; see RFC 3977
group-char     = ALPHA / DIGIT / "-" / "+" / "_" / "."
```

In the form with an <article-number>, the URL corresponds roughly to the content of an <xref> header field as specified in [RFC5536], replacing its more general <article-locator> by the <article-number> used with the NNTP.

A <newsgroup-name> as specified in [RFC5536] consists of dot-separated components. Each component contains one or more letters, digits, "-" (hyphen-minus), "+", or "_" (underscore). These characters can be directly used in a segment of a path in an [RFC3986] URI; no percent-encoding is necessary. Example:

```
nntp://news.server.example/example.group.this/12345
```

A <wildmat-exact> newsgroup name as specified in [RFC3977] allows (in theory) any <UTF8-non-ascii> (see Section 6) and most printable US-ASCII characters, excluding "!", "*", ",", "?", "[", "\", and "]". However, [RFC5536] does not (yet) permit characters outside of

<group-char> and so, to keep the syntax simple, the additional characters are here covered by <pct-encoded> as defined in [RFC3986], since most of them have to be percent-encoded anyway (with a few exceptions, such as ":", ";", and "~"). Example:

nntp://wild.server.example/example.group.n%2Fa/12345

In the form without <article-number>, the URL identifies a single group on the specified server. This is also possible with an equivalent 'news' URL, and the latter is better supported by user agents. Example:

nntp://news.server.example/example.group.this
news://news.server.example/example.group.this

4. Syntax of 'news' URIs

A 'news' URI identifies an article by its unique "Message-ID", or it identifies a set of newsgroups. Additionally, it can specify a server; when the server is not specified, a configured default server for Netnews access is used.

```
newsURL         = "news:" [ server "/" ] ( article / newsgroups )
article         = msg-id-core                   ; see RFC 5536
```

The form identifying an <article> is the <msg-id-core> from [RFC5536]; it is a "Message-ID" without angle brackets. According to [RFC3986], characters that are in <gen-delims> (a subset of <reserved>), together with the character "%", MUST be percent-encoded (though it is not wrong to encode others). Specifically, the characters allowed in <msg-id-core> that must be encoded are

"/" "?" "#" "[" "]" and "%"

Note that an agent which seeks to interpret a 'news' URI needs to decode all percent-encoded characters before passing it on to an NNTP server to be acted upon.

Please note that "%3E" (">") is not allowed; <msg-id-core> is otherwise identical to

id-left "@" id-right

as defined in [RFC5322].

The form identifying <newsgroups> corresponds to the [RFC3977] <wildmat-pattern>, a newsgroup name with wildcards "*" and "?". Any "?" has to be percent-encoded as "%3F" in this part of a URI.

Examples (the first two are equivalent):

```
news://news.server.example/*
news://news.server.example/
news://wild.server.example/example.group.th%3Fse
news:example.group.*
news:example.group.this
```

Without wildcards, this form of the URL identifies a single group if it is not empty. User agents would typically try to present an overview of the articles available in this group, likely somehow limiting this overview to the newest unread articles up to a configured maximum.

With wildcards, user agents could try to list matching group names on the specified or default server. Some user agents support only a specific <group> without wildcards, or an optional single "*".

As noted above (Section 2.2), the presence of an "@" in a 'news' URI disambiguates <article> and <newsgroups> for URI consumers. The new <message-id> construct specified in [RFC3977] does not require an "@". Since [RFC0822], the "Message-ID" syntax has been closely related to the syntax of mail addresses with an "@" separating left-hand side (local part of addresses, unique part of message identifiers) and right-hand side (domain part), and this memo sticks to the known [RFC1738] practice.

5. Acknowledgments

Henry Spencer was the driving force to adopt MIME in Netnews; he registered the MIME 'message/external-body' access type 'news-message-ID', discussed below (Section 8.2), in 1993 as recalled in "Son-of-1036" [RFC1849].

"The 'news' URL scheme" [GILMAN], by Alfred S. Gilman (March 1998), introduced additions to the original [RFC1738] 'news' URI scheme. Some of these ideas are now widely supported and reflected by the revised 'news' URI scheme specified here.

Thanks to Alfred Hoenes, Charles Lindsey, Clive Feather, Chris Newman, Ken Murchinson, Kjetil T. Homme, Lars Magne Ingebrigtsen, Martin Duerst, Matt Seitz, Nicolas Krebs, Paul Hoffman, Pasi Eronen, Roy T. Fielding, Russ Allbery, Stephane Bortzmeyer, and Tom Petch for their feedback, contributions, or encouragement.

Bill Fenner's _xml2rfc validator_ and _ABNF checker_ were a great help in the creation of (not only) this memo. The same goes for various great _IETF tools_ written by Henrik Levkowetz.

6. Internationalization Considerations

The URI schemes were updated to support percent-encoded UTF-8 characters in NNTP newsgroup names as specified in [RFC3977] and [RFC3987].

The Netnews Article Format in [RFC5536] does not yet allow UTF-8 in <newsgroup-name>s; therefore, well-known Unicode and UTF-8 security considerations are not listed below. For an overview, see [UTR36] and [RFC3629].

The work on Email Address Internationalization (EAI), started in [RFC4952], is not expected to change the syntax of a "Message-ID".

7. Security Considerations

There are many security considerations for URI schemes discussed in [RFC3986]. The NNTP protocol may use passwords in the clear for authentication or offer no privacy, both of which are considered extremely unsafe in current practice. Alternatives and further security considerations with respect to the NNTP are discussed in [RFC4642] and [RFC4643].

The syntax for the 'news' and 'nntp' URI schemes contains the general <authority> construct with an optional <userinfo> defined in [RFC3986]. As noted in [RFC3986], the "user:password" form of a <userinfo> is deprecated.

Articles on NNTP servers typically expire after some time. After that time, corresponding 'news' and 'nntp' URLs may not work anymore depending on the server. While a "Message-ID" is supposed to be worldwide unique forever, the NNTP protocol does not guarantee this. Under various conditions depending on the servers, the same "Message-ID" could be used for different articles, and attackers could try to distribute malicious content for known 'news' or 'nntp' URLs.

If a URI does not match the generic syntax in [RFC3986], it is invalid, and broken URIs can cause havoc. Compare [RFC5064] for similar security considerations.

8. IANA Considerations

The IANA registry of URI schemes has been updated to point to this memo instead of [RFC1738] for the 'news' and 'nntp' URI schemes.

8.1. 'snews' URIs

This section contains the [RFC4395] template for the registration of the historical 'snews' scheme specified in [GILMAN].

URI scheme name: snews

Status: historical

URI scheme syntax: Same as for 'news' (Section 4)

URI scheme semantics:
Syntactically equivalent to 'news', but using NNTP over SSL/TLS (SSL/TLS with security layer is negotiated immediately after establishing the TCP connection) with a default port of 563, registered as "nntps"

Encoding considerations:
Same as for 'news' (Section 6)

Applications/protocols that use this URI scheme name:
For some user agents, 'snews' URLs trigger the use of "nntps" instead of NNTP for their access on Netnews

Interoperability considerations:
This URI scheme was not widely deployed; its further use is deprecated in favor of ordinary 'news' URLs in conjunction with NNTP servers supporting [RFC4642]

Security considerations:
See [RFC4642]; the use of a dedicated port for secure variants of a protocol was discouraged in [RFC2595]

Contact: <mailto:uri@w3.org> (URI mailing list)

Change controller: IETF

References: RFC 5538, [RFC4642], [GILMAN]

8.2. 'news-message-ID' Access Type

The MIME 'news-message-ID' access type was erroneously listed as a subtype. IANA has removed 'news-message-ID' from the application subtype registry, and has added it to the access types registry defined in [RFC4289].

[RFC4289] requires an RFC (preferably on the Standards Track) for the access types registry. To provide a definition meeting this requirement, the following paragraph is reproduced verbatim from [RFC1849]:

> NOTE: In the specific case where it is desired to essentially make another article PART of the current one, e.g., for annotation of the other article, MIME's "message/external-body" convention can be used to do so without actual inclusion. "news-message-ID" was registered as a standard external-body access method, with a mandatory NAME parameter giving the message ID and an optional SITE parameter suggesting an NNTP site that might have the article available (if it is not available locally), by IANA 22 June 1993.

Please note that 'news' URLs offer a very similar and (today) more common way to access articles by their Message-ID; compare [RFC2017].

9. References

9.1. Normative References

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

[RFC3977] Feather, C., "Network News Transfer Protocol (NNTP)", RFC 3977, October 2006.

[RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, RFC 3986, January 2005.

[RFC5234] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234, January 2008.

[RFC5536] Murchison, K., Lindsey, C., and D. Kohn, "Netnews Article Format", RFC 5536, November 2009.

9.2. Informative References

[GILMAN] Gilman, A., "The 'news' URL scheme", Work in Progress, March 1998.

[POSIX] Institute of Electrical and Electronics Engineers, "The Open Group Base Specifications Issue 6", IEEE Standard 1003.1, 2004 edition.

[RFC0822] Crocker, D., "Standard for the format of ARPA Internet text messages", STD 11, RFC 822, August 1982.

[RFC1738] Berners-Lee, T., Masinter, L., and M. McCahill, "Uniform Resource Locators (URL)", RFC 1738, December 1994.

[RFC1849] Spencer, H., ""Son of 1036": News Article Format and Transmission", RFC 1849, March 2010.

[RFC2017] Freed, N. and K. Moore, "Definition of the URL MIME External-Body Access-Type", RFC 2017, October 1996.

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

[RFC2595] Newman, C., "Using TLS with IMAP, POP3 and ACAP", RFC 2595, June 1999.

[RFC2980] Barber, S., "Common NNTP Extensions", RFC 2980, October 2000.

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

[RFC3987] Duerst, M. and M. Suignard, "Internationalized Resource Identifiers (IRIs)", RFC 3987, January 2005.

[RFC4156] Hoffman, P., "The wais URI Scheme", RFC 4156, August 2005.

[RFC4157] Hoffman, P., "The prospero URI Scheme", RFC 4157, August 2005.

[RFC4248] Hoffman, P., "The telnet URI Scheme", RFC 4248, October 2005.

[RFC4266] Hoffman, P., "The gopher URI Scheme", RFC 4266, November 2005.

[RFC4289] Freed, N. and J. Klensin, "Multipurpose Internet Mail Extensions (MIME) Part Four: Registration Procedures", BCP 13, RFC 4289, December 2005.

[RFC4395] Hansen, T., Hardie, T., and L. Masinter, "Guidelines and Registration Procedures for New URI Schemes", BCP 35, RFC 4395, February 2006.

[RFC4642] Murchison, K., Vinocur, J., and C. Newman, "Using Transport Layer Security (TLS) with Network News Transfer Protocol (NNTP)", RFC 4642, October 2006.

[RFC4643] Vinocur, J. and K. Murchison, "Network News Transfer Protocol (NNTP) Extension for Authentication", RFC 4643, October 2006.

[RFC4952] Klensin, J. and Y. Ko, "Overview and Framework for Internationalized Email", RFC 4952, July 2007.

[RFC5064] Duerst, M., "The Archived-At Message Header Field", RFC 5064, December 2007.

[RFC5322] Resnick, P., Ed., "Internet Message Format", RFC 5322, October 2008.

[UTR36] Davis, M. and M. Suignard, "Unicode Security Considerations", Unicode Technical Reports #36, August 2006.

Appendix A. Collected ABNF

In addition to the syntax given above, this appendix also lists the sources of terms used in comments and the prose:

```
nntpURL         = "nntp:" server "/" group [ "/" article-number ]
server          = "//" authority               ; see RFC 3986
group           = 1*( group-char / pct-encoded )
article-number  = 1*16DIGIT                    ; see RFC 3977
group-char      = ALPHA / DIGIT / "-" / "+" / "_" / "."

newsURL         = "news:" [ server "/" ] ( article / newsgroups )
article         = msg-id-core                  ; see RFC 5536
newsgroups      = *( group-char / pct-encoded / "*" )

authority       = <see RFC 3986 Section 3.2>
host            = <see RFC 3986 Section 3.2.2>
pct-encoded     = <see RFC 3986 Section 2.1>
port            = <see RFC 3986 Section 3.2.3>
gen-delims      = <see RFC 3986 Section 2.2>
msg-id-core     = <see RFC 5536 Section 3.1.3>
reserved        = <see RFC 5536 Section 2.2>
userinfo        = <see RFC 3986 Section 3.2.1>

message-id      = <see RFC 3977 Section 9.8>
UTF8-non-ascii  = <see RFC 3977 Section 9.8>
wildmat         = <see RFC 3977 Section 4.1>
wildmat-exact   = <see RFC 3977 Section 4.1>
wildmat-pattern = <see RFC 3977 Section 4.1>

ALPHA           = <see RFC 5234 Appendix B.1>
DIGIT           = <see RFC 5234 Appendix B.1>

article-locator = <see RFC 5536 Section 3.2.14>
newsgroup-name  = <see RFC 5536 Section 3.1.4>
xref            = <see RFC 5536 Section 3.2.14>
```

Appendix B. Detailed Example

Here is an example of a mail to the <mailto:tools.discuss@ietf.org> list with "Message-ID" <p0624081dc30b8699bf9b@[10.20.30.108]>.

<http://dir.gmane.org/gmane.ietf.tools> is one of the various list archives; it converts mail into Netnews articles. The header of this article contains the following fields (among others):

```
Message-ID: <p0624081dc30b8699bf9b@[10.20.30.108]>
Xref: news.gmane.org gmane.ietf.tools:742
Archived-At: <http://permalink.gmane.org/gmane.ietf.tools/742>
```

The "Xref" roughly indicates the 742nd article in newsgroup <news://news.gmane.org/gmane.ietf.tools> on this server. An 'nntp' URL might be <nntp://news.gmane.org/gmane.ietf.tools/742>. For details about the "Archived-At" URL, see [RFC5064].

The list software and list subscribers reading the list elsewhere can't predict a server-specific article number 742 in this archive. If they know this server, they can however guess the corresponding <news://news.gmane.org/p0624081dc30b8699bf9b@%5B10.20.30.108%5D> URL.

In theory, the list software could use the guessed 'news' URL in an "Archived-At" header field, but if a list tries this, it would likely use <http://mid.gmane.org/p0624081dc30b8699bf9b@%5B10.20.30.108%5D>.

Using domain literals in a "Message-ID" could cause collisions. A collision might force the mail2news gateway in this example to invent a new "Message-ID", and an attempt to guess the future URL on this server would then fail.

Author's Address

Frank Ellermann
xyzzy
Hamburg, Germany

EMail: hmdmhdfmhdjmzdtjmzdtzktdkztdjz@gmail.com
URI: http://purl.net/xyzzy/