Network Working Group
Request for Comments: 4266
Obsoletes: 1738
Category: Standards Track

P. Hoffman
VPN Consortium

November 2005

The gopher URI Scheme

Status of This Memo

Copyright Notice

Abstract

This document specifies the gopher Uniform Resource Identifier (URI) scheme that was originally specified in RFC 1738. The purpose of this document is to allow RFC 1738 to be made obsolete while keeping the information about the scheme on standards track.

1. Introduction

URIs were previously defined in RFC 2396 [RFC2396], which was updated by RFC 3986 [RFC3986]. Those documents also specify how to define schemes for URIs.

The first definition for many URI schemes appeared in RFC 1738 [RFC1738]. Because that document has been made obsolete, this document copies the gopher URI scheme from it to allow that material to remain on standards track.

2. Scheme Definition

The gopher URL scheme is used to designate Internet resources accessible using the Gopher protocol.

The base Gopher protocol is described in RFC 1436 [RFC1436] and supports items and collections of items (directories). The Gopher+ protocol is a set of upward-compatible extensions to the base Gopher protocol and is described in [Gopher+]. Gopher+ supports associating

arbitrary sets of attributes and alternate data representations with Gopher items. Gopher URLs accommodate both Gopher and Gopher+ items and item attributes.

Historical note: The Gopher protocol was widely implemented in the early 1990s, but few Gopher servers are in use today.

2.1. Gopher URL Syntax

A Gopher URL takes the form:

gopher://<host>:<port>/<gopher-path>

where <gopher-path> is one of:

<gophertype><selector>
<gophertype><selector>%09<search>
<gophertype><selector>%09<search>%09<gopher+_string>

If :<port> is omitted, the port defaults to 70. <gophertype> is a single-character field to denote the Gopher type of the resource to which the URL refers. The entire <gopher-path> may also be empty, in which case the delimiting "/" is also optional and the <gophertype> defaults to "1".

<selector> is the Gopher selector string. In the Gopher protocol, Gopher selector strings are a sequence of octets that may contain any octets except 09 hexadecimal (US-ASCII HT or tab), 0A hexadecimal (US-ASCII character LF), and 0D (US-ASCII character CR).

Gopher clients specify which item to retrieve by sending the Gopher selector string to a Gopher server.

Within the <gopher-path>, no characters are reserved.

Note that some Gopher <selector> strings begin with a copy of the <gophertype> character, in which case that character will occur twice consecutively. The Gopher selector string may be an empty string; this is how Gopher clients refer to the top-level directory on a Gopher server.

2.2. Specifying URLs for Gopher Search Engines

If the URL refers to a search to be submitted to a Gopher search engine, the selector is followed by an encoded tab (%09) and the search string. To submit a search to a Gopher search engine, the Gopher client sends the <selector> string (after decoding), a tab, and the search string to the Gopher server.

2.3. URL Syntax for Gopher+ Items

Historical note: Gopher+ was uncommon even when Gopher was popular.

URLs for Gopher+ items have a second encoded tab (%09) and a Gopher+ string. Note that in this case, the %09<search> string must be supplied, although the <search> element may be the empty string.

The <gopher+_string> is used to represent information required for retrieval of the Gopher+ item. Gopher+ items may have alternate views and arbitrary sets of attributes, and they may have electronic forms associated with them.

To retrieve the data associated with a Gopher+ URL, a client will connect to the server and send the Gopher selector, followed by a tab and the search string (which may be empty), followed by a tab and the Gopher+ commands.

2.4. Default Gopher+ Data Representation

When a Gopher server returns a directory listing to a client, the Gopher+ items are tagged with either a "+" (denoting Gopher+ items) or a "?" (denoting Gopher+ items that have a +ASK form associated with them). A Gopher URL with a Gopher+ string consisting of only a "+" refers to the default view (data representation) of the item, and a Gopher+ string containing only a "?" refers to an item with a Gopher electronic form associated with it.

2.5. Gopher+ Items with Electronic Forms

Gopher+ items that have a +ASK associated with them (i.e., Gopher+ items tagged with a "?") require the client to fetch the item's +ASK attribute to get the form definition, and then ask the user to fill out the form and return the user's responses along with the selector string to retrieve the item. Gopher+ clients know how to do this but depend on the "?" tag in the Gopher+ item description to know when to handle this case. The "?" is used in the Gopher+ string to be consistent with Gopher+ protocol's use of this symbol.

2.6. Gopher+ Item Attribute Collections

To refer to the Gopher+ attributes of an item, the Gopher URL's Gopher+ string consists of "!" or "$". "!" refers to all of a Gopher+ item's attributes. "$" refers to all the item attributes for all items in a Gopher directory.

2.7. Referring to Specific Gopher+ Attributes

To refer to specific attributes, the URL's gopher+_string is "!<attribute_name>" or "$<attribute_name>". For example, to refer to the attribute containing the abstract of an item, the gopher+_string would be "!+ABSTRACT".

To refer to several attributes, the gopher+_string consists of the attribute names separated by coded spaces. For example, "!+ABSTRACT% 20+SMELL" refers to the +ABSTRACT and +SMELL attributes of an item.

2.8. URL Syntax for Gopher+ Alternate Views

Gopher+ allows for optional alternate data representations (alternate views) of items. To retrieve a Gopher+ alternate view, a Gopher+ client sends the appropriate view and language identifier (found in the item's +VIEW attribute). To refer to a specific Gopher+ alternate view, the URL's Gopher+ string would be in the form:

```
+<view_name>%20<language_name>
```

For example, a Gopher+ string of "+application/postscript%20Es_ES" refers to the Spanish language postscript alternate view of a Gopher+ item.

2.9. URL Syntax for Gopher+ Electronic Forms

The gopher+_string for a URL that refers to an item referenced by a Gopher+ electronic form (an ASK block) filled out with specific values is a coded version of what the client sends to the server. The gopher+_string is of the form:

```
+%091%0D%0A+-1%0D%0A<ask_item1_value>%0D%0A
<ask_item2_value>%0D%0A.%0D%0A
```

To retrieve this item, the Gopher client sends the following text to the Gopher server.

```
<a_gopher_selector><tab>+<tab>1<cr><lf>
+-1<cr><lf>
<ask_item1_value><cr><lf>
<ask_item2_value><cr><lf>
.<cr><lf>
```

3. Security Considerations

There are many security considerations for URI schemes discussed in [RFC3986]. The Gopher protocol uses passwords in the clear for authentication, and offers no privacy, both of which are considered extremely unsafe in current practice.

4. Informative References

[Gopher+] Anklesaria, F., et al., "Gopher+: Upward compatible enhancements to the Internet Gopher protocol", University of Minnesota, July 1993, <ftp://boombox.micro.umn.edu/pub/gopher/gopher_protocol/Gopher+/Gopher+.txt>

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

[RFC2396] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifiers (URI): Generic Syntax", RFC 2396, August 1998.

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

[RFC1436] Anklesaria, F., McCahill, M., Lindner, P., Johnson, D., Torrey, D., and B. Albert, "The Internet Gopher Protocol (a distributed document search and retrieval protocol)", RFC 1436, March 1993.

Author's Address

Paul Hoffman
VPN Consortium
127 Segre Place
Santa Cruz, CA 95060
US

EMail: paul.hoffman@vpnc.org

Full Copyright Statement

Intellectual Property

Acknowledgement

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