Network Working Group
Internet-Draft
Expires: July 2, 2005

P. Hoffman
VPN Consortium
January 2005

The file URI Scheme
draft-hoffman-file-uri-03.txt

Status of this Memo

Copyright Notice

Abstract

This document specifies the file 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 draft-fielding-uri-rfc2396bis [2396bis]. 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 file URI scheme from it to allow that material to remain on standards track.

2. Scheme Definition

The file URL scheme is used to designate files accessible on a particular host computer. This scheme, unlike most other URL schemes, does not designate a resource that is universally accessible over the Internet.

The file URL scheme has historically had little or no interoperability between platforms. Further, implementers on a single platform have often disagreed on the syntax to use for a particular filesystem. This document does not try to resolve those problems, only to show what has been commonly seen in use on the Internet.

Note that the file: and ftp: URIs are not the same, even when the target of the ftp: URI is the local host.

A file URL takes the form:

file://<host>/<path>

where <host> is the fully qualified domain name of the system on which the <path> is accessible, and <path> is a hierarchical directory path of the form <directory>/<directory>/.../<name>.

As a special case, <host> can be the string "localhost" or the empty string; this is interpreted as "the machine from which the URL is being interpreted". However, this part of the syntax has been ignored on many systems. That is, for some systems, the following are considered equal, while on others they are not:

```
file://localhost/path/to/file.txt
file:///path/to/file.txt
```

Some systems allow URLs to point to directories. In this case, there is usually (but not always) a terminating "/" character, such as in:

file://usr/local/bin/

On systems running some versions of Microsoft Windows, the local

drive specification is sometimes preceded by a "/" character. Thus, for a file called "example.ini" in the "windows" directory on the "c:" drive, the URL might be:

file:///c:/windows/example.ini

For Windows shares, there is an additional "/" prepended to the name. Thus, the file "example.doc" on the shared directory "department" would have the URL:

file:////department/example.doc

The file URL scheme is unusual in that it does not specify an Internet protocol or access method for such files; as such, its utility in network protocols between hosts is limited.

3. Implementation Notes

3.1 Hierarchical Structure

Most implementations of the file URI scheme do a reasonable job of mapping the hierarchical part of a directory structure into the "/" delimited hierarchy of the URI syntax, independent of what the native platform delimiter is.

For example, on Windows platforms, it is typical that the file system presents backslash "\" as the file delimeter for file names, yet the URI's forward slash "/" can be used in file: URIs. Similarly, on (some) Macintosh OS versions, at least in some contexts, the colon (":") is used as the delimiter in the native presentation of file path names. Unix systems natively use the same forward slash "/" delimiter for hierarchy, so there is a closer mapping between file paths and native path names.

3.2 Drives, drive letters, mount points, file system root

There is considerable difference, in practice, for handling of the syntax for the "top" of the hierarchy. The file URI syntax provides one simple place for designating the root of the file hierachy, and implementations have diverged, even on the same platform, sometimes even within a single application.

For example, DOS- and Windows-based systems support the notion of a "drive letter", a single character which represents a (virtual) drive, mount point, or device. Native representations of file paths start with the drive letter, a colon, and then the path; e.g., "c:\tmp\test.txt".

Drive letters are mapped into the top of a file URI in various ways, depending on the implementation; some applications substitute vertical bar ("|") for the colon after the drive letter, yielding "file:///c|/tmp/test.txt". In some cases, the colon is left unchanged, as in "file:///c:/tmp/test.txt". In other cases, the colon is simply omitted, as in "file:///c/tmp/test.txt".

3.3 Use of hostname and host name checking

The file URI specification calls for using the actual host name as the name authority and allowing it to be omitted. This practice is rarely followed, and frequently is not checked. Some applications generate URIs with no authority component at all, such as "file:/this/is/the/path".

3.4 Character sets and encodings

Local file systems sometimes use many different encodings for representing file names. For interoperability sake, it would be preferable for file: URI libraries to translate the native character encoding for file names to and from Unicode.

4. Security Considerations

There are many security considerations for URI schemes discussed in [2396bis].

File access and the granting of privileges for specific operations are complex topics, and the use of file: URIs can complicate the security model in effect for file privileges. Under no circumstance should software using file: URIs grant greater access than would be available for other file access methods.

5 Informative References

[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.

[2396bis] Berners-Lee, T., Fielding, R. and L. Masinter, "Uniform Resource Identifiers (URI): Generic Syntax", work in progress, draft-fielding-uri-rfc2396bis-nn.txt.

Author's Address

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

EMail: paul.hoffman@vpnc.org

Intellectual Property Statement

Disclaimer of Validity

Copyright Statement

Acknowledgment

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