Network Working Group
Request for Comments: 1736
Category: Informational

J. Kunze
IS&T, UC Berkeley

February 1995

Functional Recommendations for Internet Resource Locators

Status of this Memo

This memo provides information for the Internet community. This memo does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

1. Introduction

This document specifies a minimum set of requirements for Internet resource locators, which convey location and access information for resources. Typical examples of resources include network accessible documents, WAIS databases, FTP servers, and Telnet destinations.

Locators may apply to resources that are not always or not ever network accessible. Examples of the latter include human beings and physical objects that have no electronic instantiation (that is, objects without an existence completely defined by digital objects such as disk files).

A resource locator is a kind of resource identifier. Other kinds of resource identifiers allow names and descriptions to be associated with resources. A resource name is intended to provide a stable handle to refer to a resource long after the resource itself has moved or perhaps gone out of existence. A resource description comprises a body of meta-information to assist resource search and selection.

In this document, an Internet resource locator is a locator defined by an Internet resource location standard. A resource location standard in conjunction with resource description and resource naming standards specifies a comprehensive infrastructure for network based information dissemination. Mechanisms for mapping between locators, names, and descriptive identifiers are beyond the scope of this document.

2. Overview of Problem

Network-based information resource providers require a method of describing the location of and access to their resources. Information systems users require a method whereby client software can interpret resource access and location descriptions on their

behalf in a relatively transparent way. Without such a method, transparent and widely distributed, open information access on the Internet would be difficult if not impossible.

2.1 Defining the General Resource Locator

The requirements listed in this document impose restrictions on the general resource locator. To better understand what the Internet resource locator is, the following general locator definition provides some contrast.

> Definition: A general resource locator is an object that describes the location of a resource.

This definition deliberately allows many degrees of freedom in order to contain the furthest reaches of the wide-ranging debate on resource location standards. Vast as it is, this problem space is a useful backdrop for discussion of the requirements (later) that generate a smaller, more manageable problem space. A resource location standard shrinks the space again by applying additional requirements.

Consider the definition in four parts: (1) A general resource locator is an object (2) that describes (3) the location of (4) a resource.

2.1.1. A general resource locator is an object...

The object could be a complex data structure. It could be a contiguous sequence of bytes. It could be a pair of latitude-longitude coordinates, or a three-color road map printed on paper. It could be a sequence of characters that are capable of being printed on paper.

2.1.2. ...that describes

In the fully general case, there are many ways that a resource locator could describe the location. It could employ a graphical or natural language description. It could be heavily encoded or compressed. It could be lightly encoded and readily understandable by human beings. The description could be a multi-level hierarchy with common semantics at each level. It could be a multi-level hierarchy with common semantics at only the first two levels, where semantics below the second level depend on the value given at the first level. These are just a few possibilities.

2.1.3. ...the location of

A resource locator describes a location but never guarantees that access may be established. While access is often desired when clients follow location instructions given in a conformant resource locator, the resource need not exist any longer or need not exist yet. Indeed it may never exist, even though the locator continues to describe a location where a resource might exist (e.g., it might be used as a placeholder with resource availability contingent upon an event such as a payment).

Furthermore, the nature of certain potential resources, especially animate beings or physical objects with no electronic instantiation, makes network access meaningless in some cases; such resources have locators that would imply non-networked access, but again, access is not guaranteed.

2.1.4. ...a resource.

A resource can be many things. Besides the non-networked or non-electronic resources just mentioned, familiar examples are an electronic document, an image, a server (e.g., FTP, Gopher, Telnet, HTTP), or a collection of items (e.g., Gopher menu, FTP directory, HTML page). Other examples accompany multi-function protocols such as Z39.50, which can perform single round trip network access, session-oriented search refinement, and index browsing.

2.2 Producers and Interpreters of Resource Locators

Central to the discussion of locator requirements is the issue of parsability. This is the ability of an agent to recognize or understand a locator in whole or in part. Discussion may be assisted by clearly distinguishing the two main actions associated with locators.

Resource locators are both produced and interpreted. Producers are bound by the resource location standards that are in turn bound by requirements listed in this document. Interpreters of locators are not bound by the requirements; they are beneficiaries of them.

2.2.1 Resource Locator Interpreters

A resource locator is interpreted by interpreting agents, which in this document are simply called interpreters. Interpreters may be either human beings or software. Along the way to establishing access based on information in a locator, one or more interpreters may be employed. Some examples of multiple interpreters processing a single locator illustrate the concept that a resource locator may be

understandable only in part by each of several interpreters, but understandable in its entirety by a combination of interpreters.

In the first example, a software interpreter recognizes enough of a locator to understand to which external agent it needs to forward it. Here, the external agent might be a user and the locator a library call number; the software forwards the locator simply by displaying it. The agent might be a network software layer specializing in a particular communications protocol; once the service is recognized, the locator is forwarded to it along with an access request.

In another example, a human interpreter might also recognize enough of a locator to understand where to forward it. Here, the person might be a user who recognizes a library call number as such but who does not understand the location information encoded in it; the person forwards it to a library employee (an external agent) who knows how to establish access to the library resource.

A prerequisite to interpreting a locator is understanding when an object in question actually is a locator, or contains one or more locators. Some constrained environments make this question easy to answer, for example, within HTML anchors or Gopher menu items. Less constrained environments, such as within running text, make it more difficult to answer without well-defined assumptions. A resource location standard needs to make any such assumptions explicit.

2.2.2 Resource Locator Producers

Resource locators are produced in many ways, often by an agent that also interprets them. The provider of a resource may produce a locator for it, leaving the locator in places where it is intended to be discovered, such as an HTML page, a Gopher menu, or an announcement to an e-mail list.

Non-providers of resources can be major producers of locators; for example, WWW client software produces locators by translating foreign resource locators (e.g., Gopher menu items) to its own format. Some locator databases (e.g., Archie) have been maintained by automated processes that produce locators for hundreds of thousands of FTP resources that they "discover" on the Internet.

Users are major producers of resource locators. A user constructing one to share with others is responsible for conformance with locator standards. Sometimes a user composes a resource locator based on an educated guess and submits it to client software with the intent of establishing access. Such a user is a producer in a sense, but if the locator is purely for personal consumption the user is not bound by the requirements. In fact, some client software may offer as a

service to translate abbreviated, non-conformant locators entered by users into successful access instructions or into conformant locators (e.g., by adding a domain name to an unqualified hostname)

2.3 Uniqueness of Resource Locators

The topic of a "uniqueness" requirement for resource locators has been discussed a great deal. This document considers the following aspects of uniqueness, but deliberately rejects them as requirements. It is incumbent upon a resource location standard that takes on this topic to be clear about which aspects it addresses.

2.3.1. Uniqueness and Multiple Copies of a Resource

A uniqueness requirement might dictate that no identical copies of a resource may exist. This document makes no such requirement.

2.3.2. Uniqueness and Deterministic Access

A uniqueness requirement might dictate that the same resource accessed in one attempt will also be the result of any other successful attempt. This document makes no such requirement, nor does it define "sameness". It is inappropriate for a resource location standard to define "sameness" among resources.

2.3.3. Uniqueness and Multiple Locators

A uniqueness requirement might dictate that a resource have no more than one locator unless all such locators be the same. This document makes no such requirement, nor does it define "sameness" among locators (which a standard might do using, for example, canonicalization rules).

2.3.4. Uniqueness, Ambiguity, and Multiple Objects per Access

A uniqueness requirement might dictate that a resource locator identify exactly one object as opposed to several objects. This document makes no general definition of what constitutes one object, several objects, or one object consisting of several objects.

3. Resource Access and Availability

A locator never guarantees access, but establishing access is by far the most important intended application of a resource locator. While it is considered ungracious to advertize a locator for a resource that will never be accessible (whether a "networkable" resource or not), it is normal for resource access to fail at a rate that increases with the age of the locator used.

Resource access can fail for many reasons. Providers fundamentally affect accessibility by moving, replacing, or deleting resources over time. The frequency of such changes depends on the nature of the resource and provider service practices, among other things. A locator that conforms to a location standard but fails for one of these reasons is called "invalid" for the purposes of this document; the term invalid locator does not apply to malformed or non-conformant locators. Resource naming standards address the problem of invalid locators.

Ordinary provider support policies may cause resources to be inaccessible during predictable time periods (e.g., certain hours of the day, or days of the year), or during periods of heavy system loading. Rights clearance restrictions impossible to express in a locator also affect accessibility for certain user populations. Heavy network load can also prevent access. In such situations, this document calls a resource "unavailable". A locator can both be valid and identify a resource that is unavailable. Resource description standards address, among other things, some aspects of resource availability.

In general, the probability with which a given resource locator leads to successful access decreases over time, and depends on conditions such as the nature of the resource, support policies of the provider, and loading of the network.

4. Requirements List for Internet Resource Locators

This list of requirements is applied to the set of general locators defined in section 2.1. The resulting subset, called Internet locators in this document, is suitable for further refinement by an Internet resource location standard. Some requirements concern locator encoding while others concern locator function.

One requirement from the original draft list was dropped after extensive discussion revealed it to be impractical to meet. It stated that with a high degree of reliability, software can recognize Internet locators in certain relatively unstructured environments, such as within running ASCII text.

4.1 Locators are transient.

The probability with which a given Internet resource locator leads to successful access decreases over time. More stable resource identifier schemes are addressed in resource naming standards and are outside the scope of a resource location standard.

4.2 Locators have global scope.

The name space of resource locators includes the entire world. The probability of successful access using an Internet locator depends in no way, modulo resource availability, on the geographical or Internet location of the client.

4.3 Locators are parsable.

Internet locators can be broken down into complete constituent parts sufficient for interpreters (software or human) to attempt access if desired. While these requirements do not bind interpreters, three points bear emphasizing:

4.3.1 A given kind of locator may still be parsable even if a given interpreter cannot parse it.

4.3.2 Parsable by users does not imply readily parsable by untrained users.

4.3.3 A given locator need not be completely parsable by any one interpreter as long as a combination of interpreters can parse it completely.

4.4 Locators can be readily distinguished from naming and descriptive identifiers that may occupy the same name space.

During a transition period (of possibly indefinite length), other kinds of resource identifier are expected to co-exist in data structures along with Internet locators.

4.5 Locators are "transport-friendly".

Internet locators can be transmitted from user to user (e.g, via e-mail) across Internet standard communications protocols without loss or corruption of information.

4.6 Locators are human transcribable.

Users can copy Internet locators from one medium to another (such as voice to paper, or paper to keyboard) without loss or corruption of information. This process is not required to be comfortable.

4.7 An Internet locator consists of a service and an opaque parameter package.

The parameter package has meaning only to the service with which it is paired, where a service is an abstract access method. An abstract access method might be a software tool, an institution, or a network protocol. The parameter package might be service-specific access instructions. In order to protect creative development of new services, there is an extensible class of services for which no parameter package semantics common across services may be assumed.

4.8 The set of services is extensible.

New services can be added over time.

4.9 Locators contain no information about the resource other than that required by the access mechanism.

The purpose of an Internet locator is only to describe the location of a resource, not other properties such as its type, size, modification date, etc. These and other properties belong in a resource description standard.

5. Security Considerations

While the requirements have no direct security implications, applications based on standards that fulfill them may need to consider two potential vulnerabilities. First, because locators are transient, a client using an invalid locator might unwittingly gain access to a resource that was not the intended target. For example, when a hostname becomes unregistered for a period of time and then re-registered, a locator that was no longer valid during that period might once again lead to a resource, but perhaps to one that only pretends to be the original resource.

Second, because a locator consists of a service and a parameter package, potentially enormous processing freedom is allowed, depending on the individual service. A server is vulnerable unless it suitably restricts its input parameters. For example, a server that advertizes locators for certain local filesystem objects may inadvertently open a door through which other filesystem objects can be accessed.

A client is also vulnerable unless it understands the limitations of the service it is using. For example, a client trusting a locator obtained from an uncertain source might inadvertently trigger a mechanism that applies charges to a user account. Having a clear definition of service limitations could help alleviate some of these

concerns.

For services that by nature offer a great deal of user freedom (remote login for example), the pre-specification of user commands within a locator presents vulnerabilities. With careful command screening, the deleterious effects of unknowingly executing (at the client or server) an embedded command such as "rm -fr *" can be avoided.

6. Conclusion

Resource location standards, which define Internet resource locators, give providers the means to describe access information for their resources. They give client developers the ability to access disparate resources while hiding access details from users.

Several minimum requirements distinguish an Internet locator from a general locator. Internet resource locators are impermanent handles sufficiently qualified for resource access not to depend in general on client location. Locators can be recognized and parsed, and can be transmitted unscathed through a variety of human and Internet communication mechanisms.

An Internet resource locator consists of a service and access parameters meaningful to that service. The form of the locator does not discourage the addition of new services or the migration to other resource identifiers. A clean distinction between resource location, resource naming, and resource description standards is preserved by limiting Internet locators to no more information than what is required by an access mechanism.

7. Acknowledgements

The core requirements of this document arose from a collaboration of the following people at the November 1993 IETF meeting in Houston, Texas.

Farhad Ankelesaria, University of Minnesota
John Curran, NEARNET
Peter Deutsch, Bunyip
Alan Emtage, Bunyip
Jim Fullton, CNIDR
Kevin Gamiel, CNIDR
Joan Gargano, University of California at Davis
John Kunze, University of California at Berkeley
Clifford Lynch, University of California
Lars-Gunnar Olson, Swedish University of Agriculture
Mark McCahill, University of Minnesota

Michael Mealing, Georgia Tech
Mitra, Pandora Systems
Pete Percival, Indiana University
Margaret St. Pierre, WAIS, Inc.
Rickard Schoultz, KTH
Janet Vratny, Apple Computer Library
Chris Weider, Bunyip

8. Author's Address

John A. Kunze
Information Systems and Technology
293 Evans Hall
Berkeley, CA 94720

Phone: (510) 642-1530
Fax: (510) 643-5385
EMail: jak@violet.berkeley.edu