

Hypermedia and the Web

May 14, 2003

What is a resource? From RFC 2396:

Resource

A resource can be anything that has identity. Familiar examples include an electronic document, an image, a service (e.g., "today's weather report for Los Angeles"), and a collection of other resources. Not all resources are network "retrievable"; e.g., human beings, corporations, and bound books in a library can also be considered resources.

The resource is the conceptual mapping to an entity or set of entities, not necessarily the entity which corresponds to that mapping at any particular instance in time. Thus, a resource can remain constant even when its content—the entities to which it currently corresponds—changes over time, provided that the conceptual mapping is not changed in the process.

Key points:

- A resource isn't necessarily something that exists in a computer, or is retrievable via HTTP
- There is a distinction between the notion of resource (an abstraction), and its mapping to a concrete realization (e.g., a file)
- The mapping of resource to realization can change over time

URI, URL, URN

Uniform resource identifier

Uniform resource locator

Uniform resource name

Distinction between identifiers, locators, and names.

More generally, people building distributed systems tend to view names as being typically human readable (e.g., Empire State Building, www.cse.ucsc.edu), and identifiers as being generally machine readable, or readable only by a human specialist (e.g., N 40° 44.9127, W 73° 59.14962, or 128.114.48.30)

Locator:

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

An object: the structure of the sequence of bytes that comprise the locator

A descriptor of location: location can mean many things – lat/long coordinates, a (protocol, server address, pathname) triple. The same resource can have multiple location descriptors (ftp and http URLs)

Name:

A persistent identifier for a resource. Key idea is that a name will not change over the life of the resource, whereas a locator might. A name is not necessarily usable to retrieve the resource.

Identifier:

An identifier is an object that can act as a reference to something that has identity (i.e. a resource). In the case of URI, the object is a sequence of characters with a restricted syntax.

URI is the most general kind of identifier. A URI may also be a locator (URL), or a name (URN), or both.

Syntax of URIs:

{scheme}:{scheme-specific parts}

Permits scheme-specific parsing of URIs – applications parse the scheme, and then can use URI-specific parsers for the rest of the URI.

Important goals of the syntax:

- human transcribability
 - write it on paper
 - capable of being typed into a computer
 - something that people might want to remember
- scheme-specific parsing
- ability to represent hierarchy
- ability to have multiple independent schemes
- extensibility
- permit either centralized or decentralized control of the namespace, as desired

Given a specific URL, how does this provide all the information necessary to dereference a resource?

<http://www.soe.ucsc.edu/classes/cms183/Spring03/index.html>

Find out which scheme this is – parse out the “http”

Now can follow http specific parsing rules

Parse the domain name – www.soe.ucsc.edu

Use DNS protocol to look up the associated IP address (128.114.48.30) – convert name (DNS) to identifier (IP)

Open a TCP/IP connection to the IP address
Once established, send HTTP GET, with path component of URL
(classes/cms183/Spring03/index.html)

Weak links:

DNS was not originally designed to act as a human-accessed naming service.
Web dramatically increased number of lookups that did not resolve, increasing load on the DNS system.

Class discussion questions:

- Tim Berners-Lee originally believed that URLs would not be visible to navigational users of the WWW. Today we see them plastered on billboards, soda cans, etc. Where did Tim's assumption go wrong?
- Are HTTP URLs a centralized, or decentralized identifier scheme?
- Why is the design of identifiers important?
- Is a specific lat/long coordinate a resource? A locator? An identifier?