Uniform Resource Identifiers

Empire State Building
University of California, Irvine
http://www.ics.uci.edu/~fielding/

URI History
- Web Addresses [Berners-Lee] 1990
- Universal Document Identifiers 1992
- Universal Resource Identifiers RFC 1630
- URI: Locations RFC 1736,1738,1808
- URI: Names RFC 1737,2141
- Uniform Resource Identifiers RFC 2396

A simple and extensible means of identification
http://www.ics.uci.edu/~fielding/talks/
mailto:fielding@ics.uci.edu

Etymology of URI ... Identifier
- Establishing identity by reference
  - name, handle, moniker, location, ...
  - global scope
- Simple
  - Just a string of common characters
- Transcribable
  - bar napkins, advertisements, and e-mail
  - a sequence of characters, not coded character sets
- Usable
  - no additional entry barrier to deployment and use

Etymology of URI ... Resource
- What we want to identify: Resources
  - match the semantics of a hypermedia reference
  - ephemeral and persistent information
  - new and existing information sources
- A resource can be anything that has identity
  - a document or image
  - a service, e.g., "today's weather in Irvine"
  - a collection of other resources
  - non-networked objects (e.g., people)

More precisely, a resource is ...
- A temporally varying membership function \( M_r(t) \)
  that, for each time \( t \), maps to some set of semantically equivalent values
- Values may be resource representations or identifiers to other resources
- Can map to the empty set, allowing references to be made to a concept before any realization of that concept exists
- The resource is the conceptual mapping, not the entity that corresponds to that mapping at \( t \)

Representations of a Resource
- The Web is designed to manipulate and transfer representations of a resource
- A single resource may be associated with multiple representations (content negotiation)
- A representation is in the form of a media type
  - provides information for this resource
  - provides potential hypermedia state transitions
- Most representations can be cached
- GET URI transfers representation, not resource

http://www.ics.uci.edu/~fielding/talks/
**Etymology of URI ... Uniform**

- Uniformity allows
  - different types of resource identification within a single protocol element
  - uniform semantic interpretation of common syntactic elements; shared implementations
  - relative syntactic interpretation independent of scheme
  - extensibility for new identification schemes
  - bounds variability along common paths, making it easier to extend the use of URI to new applications

**Uniform Extensibility**

- URI scheme allows new types to be defined without affecting old uses
  - file, news, http, telnet, gopher, wais, ftp ...
  - Naturally lends itself to table-driven implementations
  - browser uses table to select handler
  - handler can be dynamically loaded/downloaded

  Unfortunately,
  - some people haven't figured that out yet

**Uniform Hierarchy**

- Scheme defines interpretation and structure
  - `<scheme>`: `<scheme-specific-part>`
  - Hierarchical when desired [Engelbart]
    - `<scheme>`: `<authority>`: `<path>`? `<query>`
    - `<scheme>`: `<segment>`
    - `<scheme>`: `<opaque_string>`
  - Name components may be meaningful
    - `http://www.ics.uci.edu/~fielding/talks/`
  - Hierarchy enables relative URI

**Relative URI**

- An identifier that obtains global scope when interpreted relative to a base URI
  - Only valid when the base URI is well-defined
  - Saves space
    - Allows document context to be partially independent of its location and accessibility
    - Document trees (groups of inter-related documents) can be moved without changing embedded URI
  - Documents can be shared by multiple access schemes and servers

**Design Failures**

- Transcribable, but not easily spoken
  - atch tee tee pee colon slash slash double-you double-you double-you dot eye see es dot you see eye dot ee dee you slash little fielding slash
  - Hierarchical path assumes only one root
  - not true for FTP resources, leading to ambiguity
  - gopher path isn't layered left-to-right
  - Reliance on DNS as only naming authority
  - vanity hostname explosion
  - flat namespace under dot.com

**Questions?**

- Places to see for more URI design information:
  - `http://www.w3.org/Addressing/Addressing.html`
  - `http://www.w3.org/DesignIssues/`

  Slides:
  - `http://www.ics.uci.edu/~fielding/talks/`