UNIVERSITY OF Southampton

The Architecture of the World Wide Web

COMP6017 Topics on Web Services

Dr Nicholas Gibbins – nmg@ecs.soton.ac.uk

2013-2014

UNIVERSITY OF Southampton

What is the Web?

The Web is a distributed information system that provides access to hypertext documents and other objects of interest

We have a general name for these objects of interest:

resources

A definition (from RFC3986)

"Familiar examples [of resources] include an electronic document, an image, a source of information with a consistent purpose (e.g., 'today's weather report for Los Angeles'), a service (e.g., an HTTP-to-SMS gateway), and a collection of other resources. A resource is not necessarily accessible via the Internet; e.g., human beings, corporations, and bound books in a library can also be resources. Likewise, abstract concepts can be resources, such as the operators and operands of a mathematical equation, the types of a relationship (e.g., 'parent' or 'employee'), or numeric values (e.g., zero, one, and infinity)."

University of Southampton

Architectural Bases of the Web

The notion of a resource is central to the architecture of the Web

We need to be able to:

- identify them
- represent them
- interact with them

UNIVERSITY OF Southampton

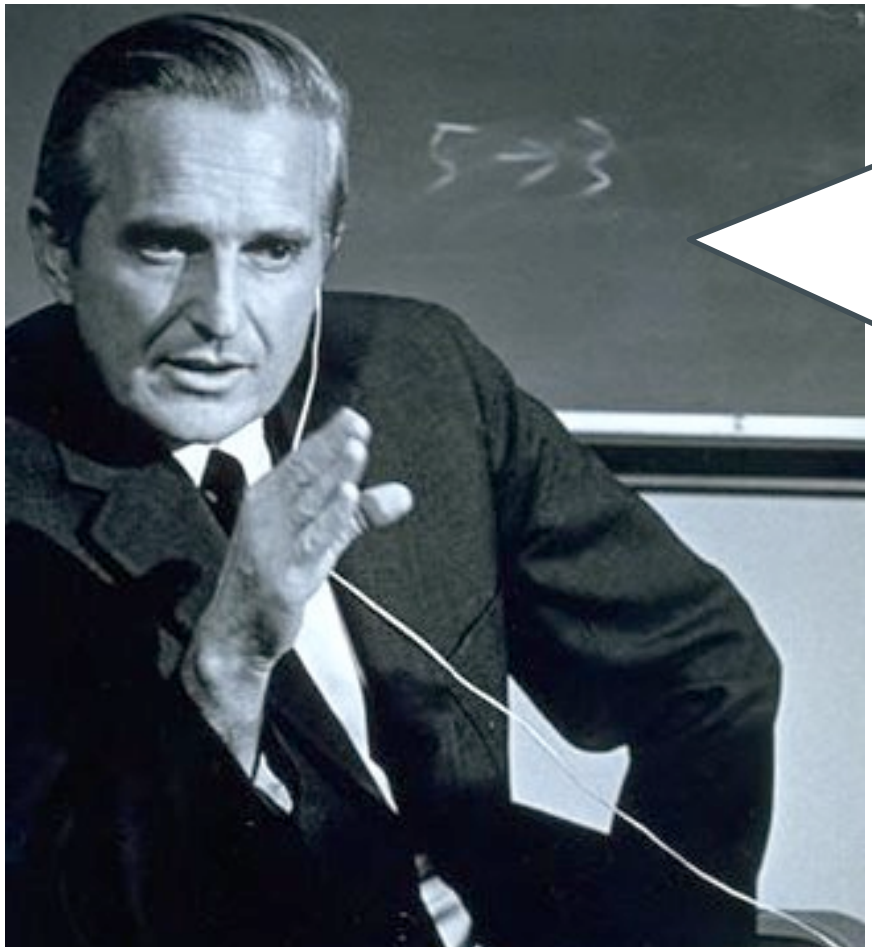
Identification

University of Southampton

Principle: Global Identifiers

Global naming leads to global network effects.

Every Object Addressable



in principle, every object that someone might validly want/need to cite should have an unambiguous address (capable of being portrayed in a manner as to be human readable and interpretable). (E.g., not acceptable to be unable to link to an object within a 'frame' or 'card.')

University of Southampton

Uniform Resource Identifiers

A “compact string of characters for identifying an abstract or physical resource”

General syntax:

<scheme>:<hierarchical part>?<query>#<fragment>

URIs, Resources and Representations

UNIVERSITY OF Southampton

URI Schemes and Examples

- http://www.example.org/
- http://www.example.org/aboutus#staff
- mailto:joe@example.org
- ftp://example.org/aDirectory/aFile
- news:comp.infosystems.www
- tel:+1-816-555-1212
- ldap://ldap.example.org/c=GB?objectClass?one
- urn:oasis:names:tc:entity:xmlns:xml:catalog

Constraint: URIs identify a single resource

Assign distinct URIs to distinct resources.

- Using the same URI to directly identify different resources produces a *URI collision*. Collision often imposes a cost in communication due to the effort required to resolve ambiguities.

Good practice: Avoiding URI aliases

A URI owner SHOULD NOT associate arbitrarily different URIs with the same resource.

- The value of a given resource can be measured by the number and value of the resources that link to it.

The Early Web

Early (pre-1991) documents refer to document naming

- "Name" and "address" used almost synonymously
- "As many protocols are currently used for information retrieval, the address must be capable of encompassing many protocols, access methods or, indeed, naming schemes"
- "A hypertext link to a document ought to be specified using the most logical name as opposed to a physical address. This is (almost) the only way of getting over the problem of documents being physically moved. As the naming scheme becomes more abstract, resolving the name becomes less of a simple look-up and more of a search."

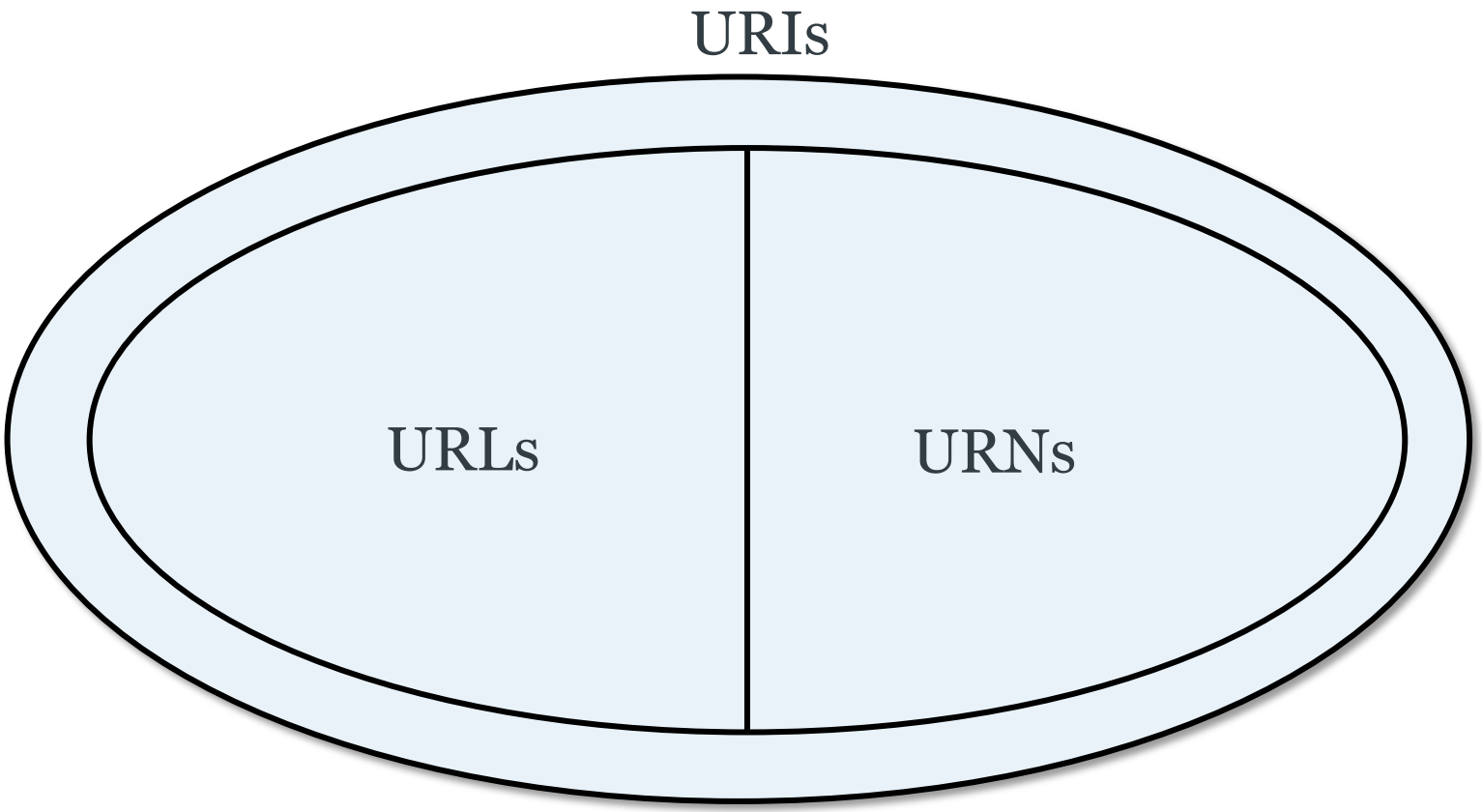
http://www.w3.org/DesignIssues/Naming

The Classical View

Early to mid-90s web specs distinguished between:

- Resource locations (URLs)
  - Often associated with network protocols (http, ftp, telnet)
- Resource names (URNs)
  - Independent of location (e.g. isbn)
- Resource identifiers (URIs)
  - Union of the above, possibly also including other classes (URCs, etc)

The Classical View



University of Southampton

The Classical View

URL resolution is (usually) well-defined

URNs don't necessarily have well-defined resolution semantics

- Resolving names depends on context
- What does resolution mean for URIs which do not refer to network resources?

University of Southampton

The Modern View

Formal URL/URN distinction is unhelpful

URL is a useful informal concept

- "a URL is a type of URI that identifies a resource via a representation of its primary access mechanism"

http://www.w3.org/TR/uri-clarification/

UNIVERSITY OF Southampton

Representation

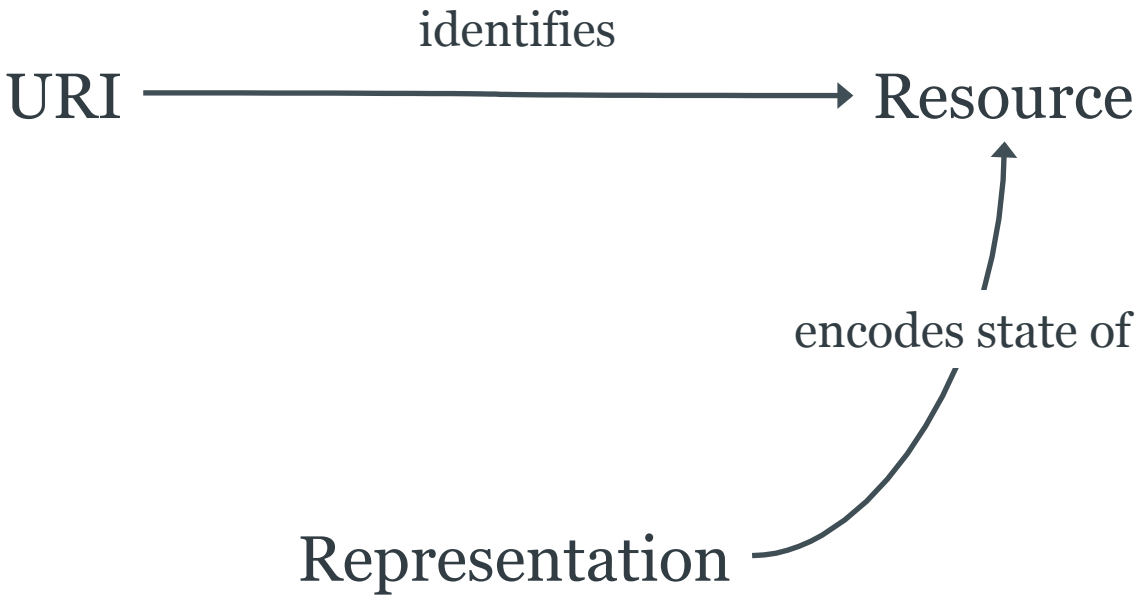
UNIVERSITY OF Southampton

Defining Representation

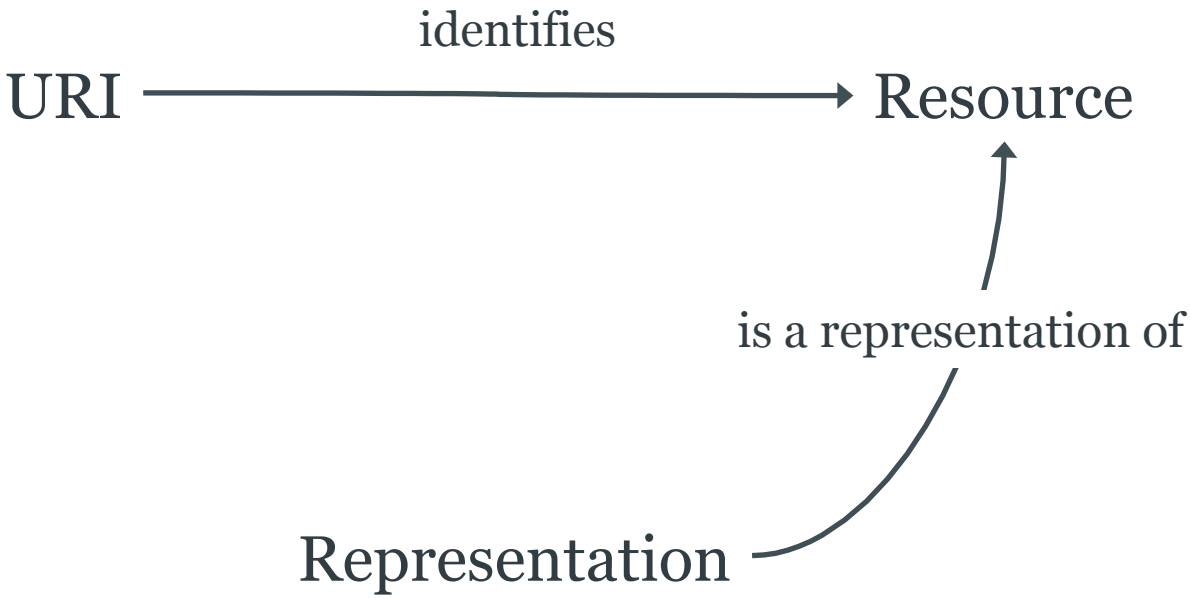
A *representation* is data that encodes information about resource state. Representations do not necessarily describe the resource, or portray a likeness of the resource, or represent the resource in other senses of the word "represent".

Representations of a resource may be sent or received using interaction protocols.

URIs, Resources and Representations



URIs, Resources and Representations



Internet Media Types

Hierarchical descriptions of data types used (originally) in email (MIME)

- Top-level types: text, image, audio, video, application (also multipart and message)
- Refinements of these top-level types:
  - text/plain, text/html, text/xml, text/csv, …
  - image/jpeg, image/gif, image/png, image/tiff, …
  - audio/mpeg, audio/ogg, …
  - video/mp4, video/quicktime, …
  - application/ecmascript, application/pdf, application/rdf+xml, …

University of Southampton

Good practice: Separation of content, presentation, interaction

A specification SHOULD allow authors to separate content from both presentation and interaction concerns.

UNIVERSITY OF Southampton

Good practice: Link identification

A specification SHOULD provide ways to identify links to other resources, including to secondary resources (via fragment identifiers).

UNIVERSITY OF Southampton

Good practice: Web linking

A specification SHOULD allow Web-wide linking, not just internal document linking.

University of Southampton

Good practice: Generic URIs

A specification SHOULD allow content authors to use URIs without constraining them to a limited set of URI schemes.

Good practice: Hypertext links

A data format SHOULD incorporate hypertext links if hypertext is the expected user interface paradigm.

UNIVERSITY OF Southampton

Interaction

Interaction

The interactions between Web agents and resources are defined in terms of protocols that control the exchange of messages

- HTTP, FTP, SOAP, NNTP, SMTP

Messages include both *data* and *metadata*

Dereferencing URIs

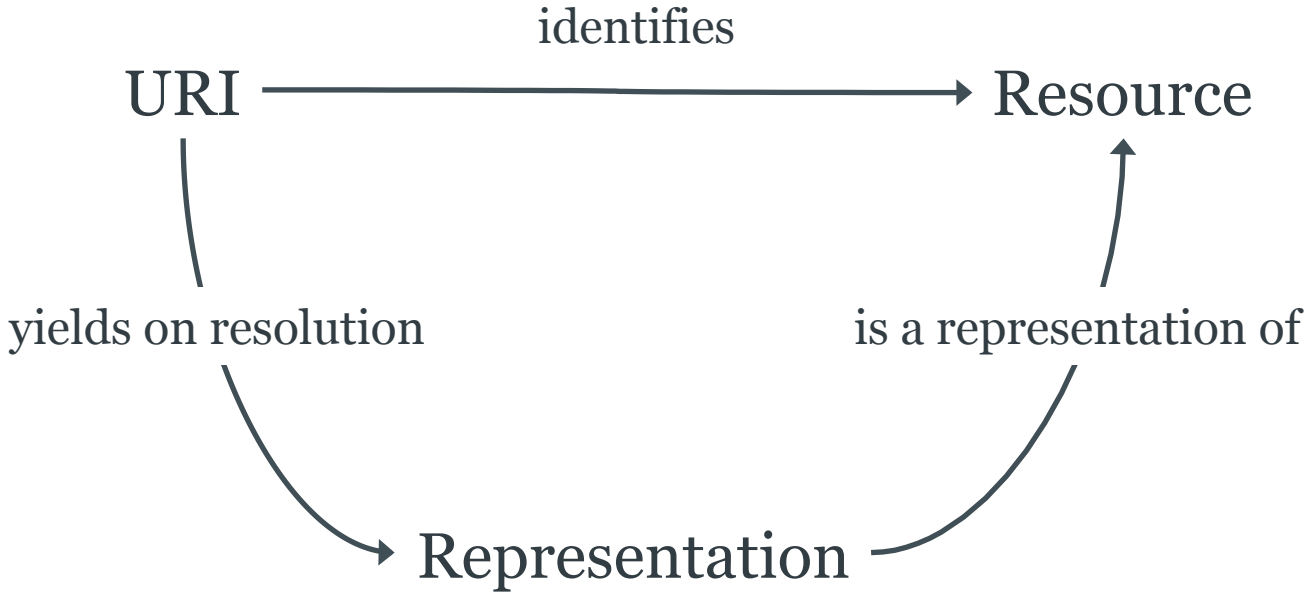
The schemas in URIs used to identify resources may indicate protocols that can be used to access those resources

- Though not always: caches, proxies, name resolution services
- Many URI schemes define a *default interaction protocol*

Resource access takes several forms:

- Retrieving a representation of the resource
- Adding or modifying a representation of the resource
- Deleting some or all representations of the resource

URIs, Resources and Representations



Good practice: Reuse representation formats

New protocols created for the Web SHOULD transmit representations as octet streams typed by Internet media types.

UNIVERSITY OF Southampton

Principle: Safe retrieval

Agents do not incur obligations by retrieving a representation.

University of Southampton

Good practice: Available representation

A URI owner SHOULD provide representations of the resource it identifies

UNIVERSITY OF Southampton

Principle: Reference does not imply dereference

An application developer or specification author SHOULD NOT require networked retrieval of representations each time they are referenced.

UNIVERSITY OF Southampton

Good practice: Consistent representation

A URI owner SHOULD provide representations of the identified resource consistently and predictably.

UNIVERSITY OF Southampton

Representational State Transfer

On REST

Architectural style that parallels and informs the design of HTTP/1.1

- Described in a thesis by Roy Fielding (Day Software, co-founder of the Apache Software Foundation, co-author of HTTP and URI RFCs)

Five key constraints:

- Client-Server
- Stateless
- Cacheable
- Layered system
- Uniform interface

Conclusion

Next lecture we'll consider how Web services are implemented within this architecture

Further Reading

Architecture of the World Wide Web, Volume One

http://www.w3.org/TR/webarch/

Architectural Styles and
the Design of Network-based Software Architectures

http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm