

Is the Web a Web of Documents or Things?

Xiaoshu Wang¹ Arlindo L. Oliveira²

¹Renaissance Computing Institute, Chapel Hill, NC, USA

²INESC-ID/IST, Lisbon, Portugal

IR-KR2009 (Pasadena, CA)

Outline

- 1 What is the Web?
 - URI, Resource, and Representation
 - Information Resource
- 2 Two Models of the Web
 - A Web of Documents
 - A Web of Things
- 3 Web as An Information System
 - Information and Meaning
 - Knowledge-Information-Data (KID)
 - Symbol-Information-Resource (SIR)
 - SIR of the Web
- 4 Pragmatics
 - Fragment Identifier
 - Metadata
 - Persistent identifier

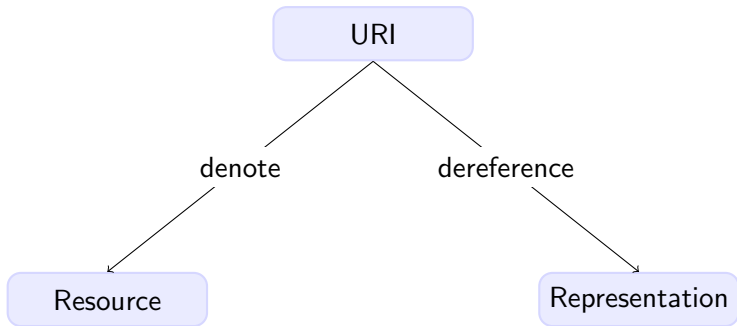
What is the Web?

The World Wide Web (WWW, or simply Web) is an information space in which the items of interest, referred to as resources, are identified by global identifiers called Uniform Resource Identifiers (URI).

The Architecture of the World Wide Web (AWWW)

<http://www.w3.org/tr/webarch>

URI, Resource, and Representation



What is Information Resource?

- **Definition:** The distinguishing characteristic of these resources is that all of their essential characteristics can be conveyed in a message. (Section 2.2 of AWWW)
- **Bylaw** (httpRange-14)
 - If 2xx response, then an information resource;
 - If 303 (See Other) response, then any resource;
 - If 4xx (error) response, then unknown.
- **Penalty**
 - What if someone does not comply to httpRange-14?

What is Information Resource?

- **Definition:** The distinguishing characteristic of these resources is that **all** of their **essential** characteristics **can** be conveyed in a message. (Section 2.2 of AWWW)
- **Bylaw** (httpRange-14)
 - If 2xx response, then an information resource;
 - If 303 (See Other) response, then any resource;
 - If 4xx (error) response, then unknown.
- **Penalty**
 - What if someone does not comply to httpRange-14?
- **Problem** The definition of IR cannot be objectively applied!

What is information?

There is one way of thinking about information. It rests on a confusion, the confusion of information with meaning. Once this distinction is clearly understood, one is free to think about information (though not meaning) as an objective commodity, something whose generation, transmission, and reception do not require or in any way presuppose interpretive processes. One is therefore given a framework for understanding how meaning can evolve, how genuine cognitive systems – those with the resources for interpretive signals, holding beliefs, and acquiring knowledge – can develop out of the lower-order, purely physical, information-processing mechanisms.

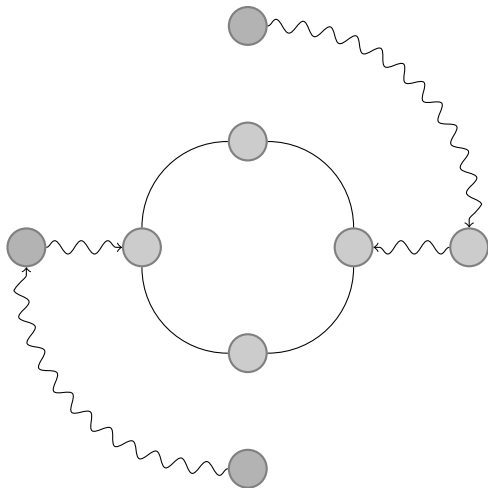
– Fred Dretske (1981) Knowledge and the Flow of Information. (p. vii)

What is information?

*There is one way of thinking about information. It rests on a confusion, **the confusion of information with meaning**. Once this distinction is clearly understood, one is free to think about **information (though not meaning) as an objective commodity**, something whose generation, transmission, and reception do not require or in any way presuppose interpretive processes. One is therefore given a framework for understanding how meaning can evolve, how genuine cognitive systems – those with the resources for interpretive signals, holding beliefs, and acquiring knowledge – can develop out of the lower-order, purely physical, information-processing mechanisms.*

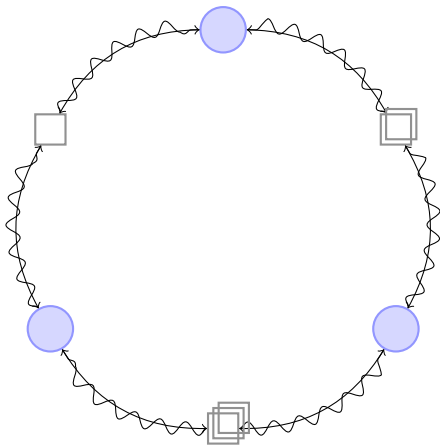
– Fred Dretske (1981) Knowledge and the Flow of Information. (p. vii)

A Web of Documents Talking about Things



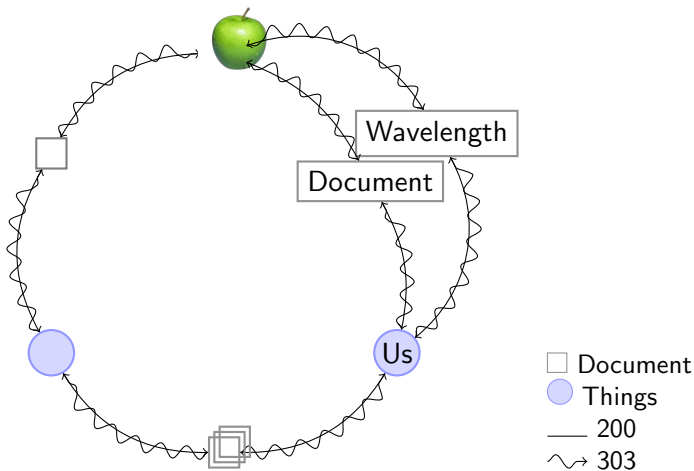
- Information Resource
- non-Information Resource
- 200
- ~ 303

A Web of Things Talking with Documents



- Document
- Things
- 200
- ~ 303

A Web of Things Talking with Documents



Information vs. Meaning

Information is objective. It must have a structure. A structureless thing cannot be information.

Meaning is subjective. It is one things being in another, i.e., how one thing stands in relation to another.

Communication The purpose of communication is to share meaning through sharing information.

Knowledge-Information-Data

Data Meaning a priori

Knowledge Meaning a posteriori

Context Direction of information flow

Resource is meaning? Yes, of course. “What is `rdfs:Resource`” is asking the same question as “What does `rdfs:Resource` mean to you?”

Symbol-Information-Resource (SIR)

SIR-Triad

Symbol Its referential range describes the expressiveness of the system

Information Defines the system's characteristics

Referent A symbol in another SIR system

Creation of Meaning

Making synonyms between symbols of different SIR system is how meaning is created.

SIR of the Web

System	Symbol	Information	Referent
Semantic Web	URN	Document	Resource
Physical Web	URL	Representation	URL Endpoint
DNS	Domain	A Record	IP Address
Internet	IP Address	Packet	Machine

Document The structure parsed from a *representation*

Representation The structure carried by a byte stream, such as the HTTP message.

Fragment Identifier Issue

What does `http://dfdf.inesc-id.pt/voc/df#byte` denote?

- A HTML element named “byte”?
- An RDF node named `http://dfdf.inesc-id.pt/voc/df#byte`?
- The concept of byte?

How to follow this advice?

“representation providers must not use content negotiation to serve representation formats that have inconsistent fragment identifier semantics” (section 3.2.2 of AWWW)

What is Metadata?

Uniformed Access to Metadata

- URI?
- Resource?
- Representation?
- Any other answer implies a fundamentally different web architecture!

Solution

All problem can and *should* be solved through content negotiation.

What is persistent identifier?

If `an:apple` is a persistent identifier, which of the following persistence does it denote?

- The binding between the Apple and natural world – i.e., a never rotten apple?
- The binding between `an:apple` and *the apple* in the nature world?
- The binding between `an:apple` and its representation? But the URI does not denote the representation.

Persistence is not a technical but a *social* one. It is effort and cannot be solved by technology no matter how we design it.

Thanks

- Arlindo L. Oliveira of INESC-ID/IST
- Ciência 2007 program of Fundação para a Ciência e Tecnologia.
- PC Chairs of IR-KR2009.