World Wide Web - History, Architecture, Protocols Web Information Systems

CS/INFO 431

January 29, 2007 Carl Lagoze - Spring 2007



#### Acknowledgments

- Erik Wilde UC Berkeley
  - http://dret.net/lectures/infosys-ws06/http

## Library as a traditional model of information management

- Functions
  - Selection
    - Collection development
    - Acquisition
  - Intermediation
    - Organization (Catalog)
    - Reference
  - Access
    - Custodianship
  - Preservation
- Characteristics
  - Professional
  - Service-oriented
  - Conservative
  - Controlled

### Standards in traditional information management

- Evolved in slow transition from elite culture to democratic culture
- Professional Culture controls adaptation
  - Shared culture through professional affiliation, ALA, IFLA
  - Shared culture through training, MLS
- Codes
  - Library Bill of Rights
  - Privacy agreements
- Intellectual Standards
  - Dewey Decimal System
  - Taxonomies LCSH, MESH
  - Cataloging Rules AACR2, Name Authorities
- Architectures
  - Machine Readable Cataloging
- Specificity, completeness, perfection

Standards in networked information management

- Roots in elite culture, revolutionary transition to democratic culture
- Complicated by profit/power potential
  - Political structures reflect this complication
- Based on code rather than human behavior
  - De facto laws based on engineering decisions
  - Tension between heuristic and algorithmic world e.g., rights management
- "Good enough" principle



Brief History and Basis of the Internet



#### The Dream Machine. J. C. R. Licklider and the Revolution That Made Computing Personal.

M. Mitchell Waldrop. author of <u>Complexity</u>.

"Waldrop's account of [Licklider's] and many others' world-transforming contributions is compelling." —John Allen Paulos, <u>The New York Times Book Review</u>



### In the beginning....



### In the beginning...







#### ARPANET



- DoD funded through leadership of Licklider
- Inspired by move from batch to timesharing
- Allowed remote login

#### Packet Switching



- Invented in early 1960's by Baran, Davies, Kleinrock
- digital, redundant, efficient, upgradeable (software)
- 1969 ARPANET first network implementation
- <u>http://en.wikipedia.org/wik</u>
  <u>i/Packet\_switching</u>

#### Packet Switching

- Network messages broken up into packets
- Each pocket has a destination address
- Pass and forward model router gets packet, examine, decides where to send next
- Message reassembled on other end

Architecture and Standards Layers

Web Semantics – Schema, RDF, OWL

Web Architecture: Protocols and Standards URI, HTTP, XML (HTML), CSS

Internet - TCP/IP, DNS

Network Hardware

Upper layers operate within constraints and opportunities of lower layers

#### Network System Design

Whenever possible, communications protocol operations should be defined to occur at the end-points of a communications system, or as close as possible to the resource being controlled.

J. Saltzer, D. Reed, D. Clark, End-to-end Arguments in System Design



#### TCP/IP Protocol Suite



- IP packet delivery
- TCP virtual circuits, packet reassembly
- ARP/RARP address resolution

#### **IP** Features

- End-to-end data transfer
- Hide lower-level heterogeneity
- Connection-less
  - Each packet routed individually
- Unreliable
  - Packets may be lost or duplicated

#### TCP Features

- Flow-controlled
  - Avoid congestion
- Reliable
  - No data lost or duplicated
- Connection-oriented
  - Virtual circuits

#### **DNS** Features

- Resolution of symbolic names to IP addresses
  - google.com -> 64.233.187.99
- Hierarchically structured
  - Root nameserver 13 exist worldwide
  - .org, .com, .uk sit under that root
  - Most root names server distributed worldwide (implemented as large clusters of machines using anycast)



#### **DNS Request Processing**



Note: in most cases cache avoids trip to higher levels Politics & business of names

- Names mean things
  - Name disputes are common
- Names have political relevance
  - ccTLDs are assigned based on the notion of what is a country
- Names can be worth \$\$\$\$
  - Cyber squatting is a popular way to make money
  - US Justice Department has record of seizing domain names that facilitate copyright infringement

#### **Internet Protocols**



Internet Issues - how to address them

- Demands of multimedia applications
- Virtual circuit reservations bandwidth and quality of service guarantees
- Real time streaming protocols
- State saving
- Political Comment
  - Increase in functionality has implications
    - Democratization of the Net
    - Privacy
    - Vulnerability

#### LAWRENCE LESSIG

AUTHOR OF

CODE AND OTHER LAWS OF CYBERSPACE

# THE FUTURE OF

THE FATE OF THE COMMONS

IN A CONNECTED WORLD

Web "Governance"



Infrastructure and Standardization

- Complex legal, economic, social, and technical process
- Wasn't invented in the information age
  - Railroad track gauge and tariffs
  - Telephone and telegraph
  - Banking
  - Power and Light
- Not for the faint-hearted

#### Internet Governance

- Internet Society (ISOC) Evolution, social & political issues
  - Mission: to assure the open development, evolution and use of the Internet for the benefit of all people throughout the world.
  - <u>http://www.isoc.org/</u>
- Internet Architecture Board (IAB) Oversees standards process
  - http://www.iab.org/
- Internet Engineering Task Force (IETF) standards development
  - <u>http://www.ietf.org/</u>
- Internet Corporation for Assigned Names and Numbers (ICANN)
  - DNS administration
  - IP # assignment
  - Protocol #'s
  - port #'s
  - Operates under contract with U.S. Department of Commerce (this is controversial!)
  - http://www.icann.org/
- World Wide Web Consortium (W3C) web standards and evolution
  - http://w3c.org

Internet Documents

- RFC's "Requests for Comments" to IETF community for information, standardization
  - <u>http://www.ietf.org/rfc.html</u>
- STD's Official IETF Internet standards
  - http://www.rfc-editor.org/rfcxx00.html
- Internet Drafts IETF working documents
  - <u>http://www.ietf.org/ID.html</u>
- W3C Reports (recommendations, drafts, notes)
  - <u>http://www.w3.org/TR/</u>

#### Well-Known Protocols

- Telnet external terminal interface, <u>RFC 854</u> (1983)
- FTP file transfer, <u>RFC 959</u> (1985)
- SMTP mail transport, <u>RFC 821 (1982)</u>
- HTTP distributed, collaborative hypermedia systems, RFC 1945 (1.0 1996), <u>RFC 2616</u> (1.1 1999)

#### Short History and Premises of the Web

- Vannevar Bush 1945 Memex
  - Associative method of information gathering, information trails
- Ted Nelson 1960's
  - "interactive and interlinked" writing, "hypertext"
  - Project Xanadu
- Hypercard 1987 Apple Computer
  - First wide-spread hypermedia system
- Tim Berner-Lee 1989, 1990 CERN
  - information sharing in a fluid context
  - "a single user-interface to many large classes of stored information such as reports, notes, data-bases, computer documentation and on-line systems help."
- W3C 1994
  - MIT
  - DARPA, ERCIM, Keio

#### Premises of the Web

- Relationships are not hierarchical
- Non-centralized management
- Anarchic Scalability
  - Clients not expected to keep track of all servers
  - No central link registry (no back links)
- Failure tolerance nodes can come and go
- Clean division of document display and format (browsers and HTML) from access (HTTP)

#### Early Depiction of Web Intererability



@ 1992 Tim Berners-Lee, Robert Cailliau, Jean-François Groff, C.E.R.N.

Basic Web Technologies

- Document layout
  - HTML  $\rightarrow$  XML
- Document formatting
  - CSS
- Document naming
  - URI's
- Document typing
  - MIME
- Document access
  - HTTP



#### **Information Management: A Proposal**

Tim Berners-Lee, CERN March 1989, May 1990

http://www.w3.org/History/1989/proposal.html



#### Web as a graph

We can call the circles nodes, and the arrows links. Suppose each node is like a small note, summary article, or comment. I'm not over concerned here with whether it has text or graphics or both. Ideally, it represents or describes one particular person or object. Examples of nodes can be

- People
- Software modules
- Groups of people
- Projects
- Concepts
- Documents
- Types of hardware
- Specific hardware objects

#### Web as a graph

The arrows which links circle A to circle B can mean, for example, that A...

- depends on B
- is part of B
- made B
- refers to B
- uses B
- is an example of B