Introduction to:
The Architecture of the World Wide Web

Noah Mendelsohn
Tufts University
Email: noah@cs.tufts.edu
Web: http://www.cs.tufts.edu/~noah
What you should get from this session

- You should understand at a high level the three pillars of Web Architecture
- You should understand what happens when a Web page is retrieved using HTTP
- You should know to refer to the TAG’s “Architecture of the World Wide Web” for more information
- You should understand the difference between open standards and open source, and why both are important to the Web
History
Early History of the Web

- Started out as a system for distributing documents written by scientists at the CERN physics lab in Switzerland – initial proposal in 1989
- A chance to realize Tim Berners-Lee’s vision: a system for integrating all the world’s information!
- August 6, 1991: announcement and early code made available along with a server you could access (Tim’s server computer, Browser screen, Line Mode Browser)
- Others start writing code that complies with Web protocols (HTTP) and document formats (HTML)
- Mosaic browser provides first widely available graphical interface
- April 1993: Tim convinces CERN to give away the Web’s technology and code
Goals and requirements for the Web

- Integrate all of the world’s online information
- **Integrate with other systems**
  - The Web is implemented on systems ranging from mainframes to traffic lights
- **Allow references (URIs) to be:**
  - Memorable
  - Conveyed in other systems (like the links in this slide show!)
  - Written “on the side of a bus”
  - Broken!
- **Explorable** – random browsing should work, and should do no harm
- **Support all users, regardless of location, spoken language or disability**
- **Extensible** to new types of content, new devices, new modalities of interaction, etc.
- **Open:** content, naming and extensions should not require concurrence of a central authority
- **Safe to use:** e.g. should not unduly compromise your privacy
- **Provide non-discriminatory access**
Web Architecture Basics
Demonstration:
http://webarch.noahdemo.com/demo1/test.html

A simple Web page retrieval
Architecting a universal Web

- Identification: URIs
- Interaction: HTTP
- Data formats: HTML, JPEG, GIF, etc.
The user clicks on a link

URI is http://webarch.noahdemo.com/demo1/test.html
The http “scheme” tells client to send HTTP GET msg

URI is http://webarch.noahdemo.com/demo1/test.html
The server is identified by DNS name in the URI

URI is http://webarch.noahdemo.com/demo1/test.html

HTTP GET

Host: webarch.noahdemo.com
The client sends an HTTP GET

URI is \texttt{http://webarch.noahdemo.com/demo1/test.html}

GET /demo1/test.html HTTP/1.0
Host: webarch.noahdemo.com
User-Agent: Noahs Demo HttpClient v1.0
Accept: */*
Accept-language: en-us
The server sends an HTTP Response

HTTP/1.1 200 OK
Date: Tue, 28 Aug 2007 01:49:33 GMT
Server: Apache
Transfer-Encoding: chunked
Content-Type: text/html

<html>
<head>
<title>Demo #1</title>
</head>
<body>
<h1>A very simple Web page</h1>
</body>
</html>
The server sends a

HTTP/1.1 200 OK
Date: Tue, 28 Aug 2007 01:49:33 GMT
Server: Apache
Transfer-Encoding: chunked
Content-Type: text/html

<html>
<head>
<title>Demo #1</title>
</head>
<body>
<h1>A very simple Web page</h1>
</body>
</html>

The “representation” returned is an HTML document
HTTP/1.1 200 OK
Date: Tue, 28 Aug 2007 01:49:33 GMT
Server: Apache
Transfer-Encoding: chunked
Content-Type: text/html

<html>
<head>
<title>Demo #1</title>
</head>
<body>
<h1>A very simple Web page</h1>
</body>
</html>
Three pillars of Web Architecture

**URI is** http://webarch.noahdemo.com/demo1/test.html

Identification with URIs

HTTP GET

HTTP RESPONSE

Host: webarch.noahdemo.com

demo1/test.html
Three pillars of Web Architecture

URI is http://webarch.noahdemo.com/demo1/test.html

HTTP GET

Interaction Using HTTP

HTTP RESPONSE

demo1/test.html

Host: webarch.noahdemo.com
Three pillars of Web Architecture

HTTP GET

HTTP/1.1 200 OK
Date: Tue, 28 Aug 2007 01:49:33 GMT
Server: Apache
Transfer-Encoding: chunked
Content-Type: text/html

<!DOCTYPE html>
<html>
<head>
<title>Demo #1</title>
</head>
<body>
<h1>A very simple Web page</h1>
</body>
</html>

Representations using media types like text/html, image/jpeg, etc

Host: webarch.noahdemo.com

demo1/test.html
Architecting a universal Web

- Identification: URIs
- Interaction: HTTP
- Data formats: HTML, JPEG, GIF, etc.

Suggested Reading:

The Architecture of the World Wide Web
http://www.w3.org/TR/webarch/
Why the Experts Knew It Wouldn’t Work
Mark Frisse, now of Vanderbilt University, was one of the reviewers of Tim Berners-Lee's original paper presenting the concepts behind the World Wide Web in 1989. Mark describes how he thought that the architecture would not scale, and that Tim's decision to allow "broken" pointers (i.e. violate bidirectional integrity) would lead to a "spaghetti bowl of gotos." Tim's paper was relegated to a poster session, which turned out to be wildly successful. Tim's decision to relax the requirement for bidirectional integrity (allowing 404 not found error) turned out to be one key feature for the success of the web.

Video:
https://archive.org/details/Munnecke-MarkFrissesMessageToTimBernersLee313
Open Standards and Open Source
Open protocol and format *standards*

**URI is** http://webarch.noahdemo.com/demo1/test.html
Open protocol and format *standards*

**URI is** http://webarch.noahdemo.com/demo1/test.html

Open Standards protocols and formats: client doesn’t see resource/server details
Open protocol and format *standards*

URI is http://webarch.noahdemo.com/demo1/test.html

HTTP GET

HTTP RESPONSE

Open Standards protocols and formats: server supports any client
Open source software

URI is http://webarch.noahdemo.com/demo1/test.html

Open Software sometimes useful for implementing servers or clients – promotes open standards protocols and formats: server supports any client