

Applications

HTTP

- The *World Wide Web* (WWW) is what many people wrongly think is the Internet
- This is a testament to the fact that the WWW was the major driving force behind the growth, commercialisation and general acceptance of the Internet
- HTTP is the protocol that drives the WWW

Applications

HTTP

- The WWW is at base very simple:
 - client makes a request
 - a server delivers some data
- So the *Hyper Text Transfer Protocol* (HTTP) started as a simple request-response protocol
- To make it easy to debug, it is human-readable

Applications

HTTP

- A request for a Web page could be

```
GET /home.html
```

```
HTTP/1.0
```

```
Host:wibble:80
```

(there is a blank line at the end of this)

- This is an example of HTTP version 1 for simplicity: generally version 1.1 is used

Applications

HTTP

- A request for a Web page could be

GET /home.html

HTTP/1.0

Host:wibble:80

- This requests the file home.html from the server wibble, connecting on port 80

Applications

HTTP

- 80 is the conventional port number for HTTP
- The blank line ends the HTTP header:
sometimes additional information follows when
the client needs to send more data to the server

Applications

```
HTTP/1.0 200 OK
Server: Wombat Server 1.0
MIME-version: 1.0
Content-type: text/html
Content-length: 81
Expires: Tue Feb 14 22:57:29
2006
```

```
<html>
<head>
<title>Example</title>
</head>
<body>
Hello world.
</body>
</html>
```

- The response contains a HTTP header followed by the data, some HTML in this case

Applications

```
HTTP/1.0 200 OK
Server: Wombat Server 1.0
MIME-version: 1.0
Content-type: text/html
Content-length: 81
Expires: Tue Feb 14 22:57:29
2006
```

```
<html>
<head>
<title>Example</title>
</head>
<body>
Hello world.
</body>
</html>
```

- The HTTP version the server uses, followed by 200, a code to indicate success
- A description of the server software
- A MIME header that encapsulates the data

Applications

```
HTTP/1.0 200 OK
Server: Wombat Server 1.0
MIME-version: 1.0
Content-type: text/html
Content-length: 81
Expires: Tue Feb 14 22:57:29
2006
```

```
<html>
<head>
<title>Example</title>
</head>
<body>
Hello world.
</body>
</html>
```

- The MIME header gives additional information, such as the kind of data, its length, and its lifetime

Applications

```
HTTP/1.0 200 OK
Server: Wombat Server 1.0
MIME-version: 1.0
Content-type: text/html
Content-length: 81
Expires: Tue Feb 14 22:57:29
2006
```

```
<html>
<head>
<title>Example</title>
</head>
<body>
Hello world.
</body>
</html>
```

- Web pages can change rapidly or slowly and the Expires date indicates how long this page should be cached by the client: it need not refetch this page until then

Applications

```
HTTP/1.0 200 OK
Server: Wombat Server 1.0
MIME-version: 1.0
Content-type: text/html
Content-length: 81
Expires: Tue Feb 14 22:57:29
2006
```

```
<html>
<head>
<title>Example</title>
</head>
<body>
Hello world.
</body>
</html>
```

- The blank lines ends the header and the data follows

Applications

HTTP

- HTTP versions before 1.1 opened a new TCP connection for each HTTP request
- A single Web page can easily require dozens of requests: one for the page and one each for each image
- This is a lots of overhead in the TCP handshake and teardown

Applications

HTTP

- Also TCP flow control does not get up to speed
- We could use T/TCP, but better is a *persistent connection*
- This is a TCP pipe that stays open and is used for more than one HTTP request
- This is more complicated, but results in better throughput

Applications

HTTP Proxy

- Another feature is *proxying*
- A host can be set up as a HTTP proxy that forwards HTTP requests for hosts that do not have a direct connection to the Internet (usually for security)
- A client sends all its request to the proxy and the proxy forwards them

Applications

HTTP Proxy

- The HTTP header contains enough information (the Host field) for the proxy server to know where to forward the request
- The proxy server then relays the results back to the original client

Applications

HTTP Proxy

- Security: administrators can focus on making the proxy server as secure as possible and the clients can be protected by a firewall
- Monitoring: the proxy server can record all the requests it passes on
- Control: the proxy server can be configured to refuse to connect to certain WWW sites

Applications

Secure HTTP

- A variant of HTTP, *Secure* HTTP (HTTPS), can be used when it is important the data can't be read by an eavesdropper as it passes by
- This is just HTTP run over an encrypting layer SSL/TLS

Applications

Secure HTTP

- In contrast, SHTTP is an extension to the HTTP protocol to include security and authentication
- HTTPS is by far the most used in the real world: make sure a page is using HTTPS whenever you send confidential data (such as credit card numbers) over the Web!

Applications

HTML and XML

- The *Hyper Text Markup Language* (HTML) is the “language of the Web” and describes how a Web page should look
- Very large and complicated language
- A “simplified” version of the more general *Standard Generalized Markup Language* (SGML)

Applications

HTML

- HTML uses *tags* like `` and `` to mark areas of text that should be display in a special way
- `` and `` indicate **boldface**
- `<hr>` is to draw a horizontal rule across the page
- And so on for a *huge* number of tags

Applications

HTML

- The main feature of HTML is the use of *anchor links*, such as

Time for `the News`

to include hypertext links to other documents

Applications

HTML

- This might appear in a browser as

Time for the News

- The browser will fetch the appropriate document when the user clicks on the link
- This is what makes it the “Web”: documents are joined to documents in an immense and complicated network

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html>
<head>
<title>Evening News</title>
</head>
<body>
<h2 align="center">News: Evening Edition</h2>
```

News for the twilight hours

```
<p>
```

Last updated: Tuesday, 31st May, 2005, 16:23 BST

Copyright: Acme News Corporation

```
</p>
```

```
<h2>World Peace Achieved</h2>
```

Officials were stunned when peace spontaneously broke out this morning.

```
<a href="http://www.acmenews.com/news/item1234.html">Link</a>
```

```
<h2>Aliens visit the Taj Mahal</h2>
```

The space travellers

from Betelgeuse continue on their world tour.

```
<a href="http://www.acmenews.com/news/item2134.html">Link</a>
```

```
</body>
```

```
</html>
```

Applications

This could be rendered as:

News: Evening Edition

News for the twilight hours

Last updated: Tuesday, 31st May, 2005, 16:23 BST

Copyright: Acme News Corporation

World Peace Achieved

Officials were stunned when peace spontaneously broke out this morning. [Link](#)

Aliens visit the Taj Mahal

The space travellers from Betelgeuse continue on their world tour. [Link](#)

Applications

XML

- What's turning out to be just as important as HTML is the *Extensible Markup Language* (XML)
- This is a markup language in the style of HTML and SGML but the data it marks up is not just text
- While HTML is concerned with the layout of a document, XML is there to describe what the data means

Applications

XML

- A simple example is *RDF Site Summary* (RSS, aka *Rich Site Summary*), a lightweight XML markup that is primarily use to provide news headlines on the Web
- RSS is part of the *Resource Description Framework* (RDF)

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<?xml-stylesheet href="http://www.acmenews.com/rss/rss.css" type="text/css"?>

<!DOCTYPE rss SYSTEM "http://www.acmenews.com/rss/rss-0.91.dtd">
<rss version="0.91">
<channel>
<title>News: Evening Edition</title>
<link>http://www.acmenews.com/home.html</link>
<description>News for the twilight hours</description>
<language>en-gb</language>
<lastBuildDate>Last updated: Tuesday, 31st May, 2005, 16:23 BST</lastBuildDate>
<copyright>Copyright: Acme News Corporation</copyright>

<docs>http://www.acmenews.com/news/</docs>

<image>
  <title>Acme News</title>
  <url>http://www.acmenews.com/pics/eve_news.gif</url>
  <link>http://www.acmenews.com/home.html</link>
</image>

<item>
  <title>World Peace Achieved</title>
  <description>Officials were stunned when peace spontaneously broke out this
morning.</description>
  <link>http://www.acmenews.com/news/item1234.html</link>
</item>

<item>
  <title>Aliens visit the Taj Mahal</title>
  <description>The space travellers from Betelgeuse continue on their world
tour.</description>
  <link>http://www.acmenews.com/news/item2134.html</link>
</item>

</channel>
</rss>
```

Applications

XML

- Associated with this is a *Document Type Definition* (DTD)
- This is a description of what are legal elements in this markup and how they can be nested
- It allows us to check if the document is syntactically correct

Applications

XML

```
<!ELEMENT item (title | link | description)*>  
<!ELEMENT title (#PCDATA)>
```

- This describes how an item element can contain a title, a link or a description element (and only these)
- A title can only contain (parsed character) data

Applications

XML

- The tags in XML describe *what is the meaning or purpose of the enclosed data*
- The tags in HTML describe *how to display the enclosed data*
- By itself, XML contains no information on how to display things
- Actually, most XML is never displayed

Applications

CSS

- In those cases we wish to display an XML document we can have a *Cascading Style Sheet* (CSS) associated with the XML document
- This describes how to display the document
- It is “cascading” as more than one style sheet can be associated with a document and priority rules describe how style are combined

Applications

CSS

```
item title {  
  display:block;  
  font-family:"Verdana",sans-serif;  
  font-size:large;  
  font-weight:bold;  
  text-align:left;  
}
```

- This says a title within an item element should be displayed in a large bold Verdana font aligned to the left on the page

Applications

CSS

```
item link {  
  display:block;  
}  
item link:before {  
  content: "Link: <";  
}  
item link:after {  
  content: ">";  
}
```

- This puts the text “Link: <” before a link inside an item and “>” after
- With this (and other) styles the XML document above might be displayed as follows

Applications

News: Evening Edition

News for the twilight hours

Last updated: Tuesday, 31st May, 2005, 16:23 BST

Copyright: Acme News Corporation

World Peace Achieved

Officials were stunned when peace spontaneously broke out this morning.

Link: <<http://www.acmenews.com/news/item1234.html>>

Aliens visit the Taj Mahal

The space travellers from Betelgeuse continue on their world tour.

Link: <<http://www.acmenews.com/news/item2134.html>>

Applications

CSS

- Closely related to CSS is the *Extensible Style Language (XSL)*
- This does the same as CSS but is written in XML, rather than a separate language
- Also, *XSL transforms* the XML document using the *XSL Transformations Language (XSLT)* into (say) HTML plus CSS that can then be displayed

Applications

XML

- The important point about XML is that it can be easily parsed to extract the information it contains
- CSS and rendering is secondary to this
- It is easy to extract the headlines in the XML document above: they are given by title elements inside items and are nowhere else

Applications

XML

- Compare this with finding the headlines in the HTML document: there no particular indication where the headlines are
- HTML is about display, not content

Applications

XML

- Finding the copyright in the XML is easy: it has its own tag
- Finding the copyright in the HTML is problematic: searching for the word “copyright” is not enough as it might be an article about copyright
- The act of going through HTML and trying to reconstruct information is called *scraping* and is never foolproof

Applications

XML

- XML is used to mark up all kinds of information
- The *Organization for the Advancement of Structured Information Systems* (OASIS) exists to publish DTDs for various e-commerce activities
- Having widely accepted standards maximised the interoperability of data exchange between institutions

Applications

XML Examples

- *Simple Object Access Protocol (SOAP)*. Part of a Web-based remote procedure call mechanism, similar in function to Sun's RPC. Elements exist for integers, floats, strings, etc.
- *Web Services Description Language (WSDL)*. To specify the location of a service and the operations the service provides, somewhat like the portmapper. Allied with SOAP

Applications

XML Examples

- *Mathematical Markup Language* (MathML). Describes mathematics in a way that it is easy to extract the meaning. MathML delivers mathematics just as HTML delivers text
- *Open Financial Exchange* (OFX). A specification for the exchange of financial data between institutions, businesses and customers

Applications

XML Examples

- *XML User-Interface Language (XUL)*. A language for describing user interfaces. For example, the layout of a browser (the “chrome”) can be described using XUL
- *Astronomical Markup Language (AML)*. For controlling astronomical instruments
- *Wireless Markup Language (WML)*. A version of HTML for low-rate telephone links.

Applications

XML Examples

- *Scalable Vector Graphics (SVG)*. Describes two dimensional vector graphics
- *MusicXML*. Represents (Western notation) musical scores, sheet music an so on
- *Voice Extensible Markup Language (VoiceXML)*. For voice user interfaces, synthesised speech, recognition and recording

Applications

XML Examples

- *DocBook*. Books, papers and presentations. Elements mark chapters, sections, titles and so on. Tools convert DocBook to HTML and more traditional typesetters such as LaTeX
- Open Document. Office applications, such as documents, spreadsheets, presentations and so on. In direct competition with Microsoft's proprietary formats for the same applications

Applications

XML Examples

- And so on for a large number of applications
- There is even XHTML, and XML markup that looks very much like HTML. It is indented that XHTML replaces HTML as then all the XML tools are then applicable to Web pages
- Anyone can make up a DTD, but its use will be fairly limited until it gets wide acceptance, which is why recognition by a standards body is important