

University of Edinburgh

M.Sc. Informatics

Computer Networking assignment 2:

Essay on the World Wide Web:

past, present and future.

Bruce J. Lamond 9653742

7th December 2001

Introduction

An ever-increasing number of millions of people around the world can testify to the fact that the World Wide Web has arrived. The last few years have seen the realisation of the potential in this information-interconnecting suite of mechanisms, particularly in the areas of science, commerce and the Media. From simple beginnings in 1990 as a network filing cabinet, the Web has grown into a colossal data repository, with the current focus being on delivery of ever more sophisticated inclusive multimedia. The future direction may well have an impact of a similar magnitude as the first decade has, with the Web evolving into a semantically charged, highly structured, ever faster organisation. In this essay I intend to discuss the Web in a basic chronology of past to future, in terms of the suite of mechanisms which underpin its operation. I conclude with an examination of its likely development into a sophisticated information grid.

Past to present

Historical

Although the World Wide Web's (hereafter referred to as 'the Web') recent historical exponential growth can readily be appreciated as phenomenal and completely surprising to the majority of the general population and most experts, it is nonetheless built on decades of deliberate scientific enterprise. To gain insight into this technology requires a brief examination of its precedents.

'Hypertext' can be understood as electronically represented and cross-referenced text, an idea flowering in 1945. Before the Web could come to fruition, the Internet itself had to develop in terms of both physical and abstract capabilities. Initially in 1972 DARPA began work which led to the creation of the Internet. With the creation of NSFNET in 1986, speeds of 56 kb/s were possible and with the subsequent upgrade to T1 in 1988 this figure increased to 1.544 Mb/s. Of course none of the data would have got anywhere without the coincident development of internet protocols, from File Transfer Protocol (FTP) and Transmission Control Protocol (TCP) in 1973/74 to the establishment of Internet Protocol (IP) and TCP (TCP/IP) as the standards for the Internet in 1982 [1].

A proposal was circulated in 1990 by Tim Berners-Lee at CERN and co-authored by Robert Cailliau, with the purpose of managing information within the disparate and diverse High-Energy Physics community over the Internet. Entitled "World Wide Web: an Information Infrastructure for High-Energy Physics", the proposal was accepted and the Web was launched within the busiest Internet node in Europe within a few months. The mechanism used a simple protocol to access remote information over a network in a Hypertext-based format common to client and server, using existing technology for display, search, and document-authoring functions. Document collections were actively maintained and made accessible through keyword searches and hyperlinks, and all of this was implemented with public-

domain software and made available free-of-charge. At the end of 1991, shortly after the Internet backbone upgrade to T3 allowed rates of 45 Mb/s, the Web was released to the rest of the World [2]. To make the Web's information resources as widely available as possible required the development of three key mechanisms: a uniform naming scheme for resource location, protocols to afford access, and a Hypertext representation for document navigation. These ideas can be followed from 1990 to the present and beyond.

Uniform Resource Identifier (URI).

Developed in 1990 and specified in 1998 in RFC 2396 [3], URI's (the familiar Uniform Resource Locator URL is the kind of URI used on the Web) are semantic string constructions which allow the location of a resource to be mapped onto an access algorithm using inter-network protocols. This is achieved by assigning an abstract and unique identification to the resource according to the following syntax:

<scheme>:<scheme-specific-part>

where <scheme> can be any of a large number of protocol or information access schemes such as FTP, Hypertext Transfer Protocol, Gopher and others. The <scheme-specific-part> form depends on the particular scheme, but those URI's involving the direct use of an IP-based protocol are composed of a double forward-slash followed by optional name and password, host name and port number, then a hierarchical pathname separated by forward slashes, similar to filesystem notation.

Hypertext Transfer Protocol (HTTP).

HTTP is a multifunctional and highly versatile application-level protocol used in distributed hypermedia information-switching systems. Originally developed in 1990 as a simple raw data transfer protocol (version 0.9), the 1994 HTTP 1.0 specification was extended to incorporate Internet Media Types (previously called MIME Types), before the present standard version 1.1 was introduced in 1997 and specified in 1999 in RFC2616 [4]. This improved version now includes consideration for hierarchical proxies (forwarding agents), improved caching (message response and directing), persistent TCP connections, and support for virtual hosts (multiple domain mapping to single IP addresses). Figure 1 shows the common implementation of HTTP at the application-level over TCP (connection-oriented) and IP (connectionless) services, although HTTP can actually be implemented on top of any IP, the criterion being reliability of the transport layer. HTTP's versatility stems from its application of a variable set of methods to identify the nature of an information request. The protocol also allows access to diverse hypermedia resources for other protocols by facilitating communication between a user agent and the relevant gateway (receiving agent) & proxy set-up. Problems associated with proliferation of Web use surfaced in the mid-nineties, ushering in the changes between versions 1.0 and 1.1. Version 1.0 was found to be inhibiting flow on the Network because it entailed establishing a separate TCP connection for each resource in a document. This involved the transmission of a three-way handshake,

client request, and server response and closure of connection (using a further two stage two-way handshake) for each inlined resource in a document. One can see how this would dramatically increase congestion with today's commonly multiple-element Web pages. Multiple connection establishment like this was also incurring time delays due to the 'slowstart' technique of TCP. Slowstart is where the protocol sends one message, doubling in size to two, four etc (on receipt of corresponding acknowledgements), until the message size is that required by the receiver. The technique is meant to ease the load on congested links but as this slow build up is used on each new connection, each resource in an HTTP 1.0 transfer is delayed [23]. The recent rapid increase in multimedia involvement in the Web has seen a corresponding increase in the number of media formats which must be supported by HTTP. Five primary media types exist - *text*, *image*, *audio*, *video*, and *application* plus three composite media types - *multipart*, *message*, and *model*. Each has a number of subtypes, for example *text/html* represents lines of text in the format of an HTML document (following section) and *image/jpeg* is an image with JPEG compressed file format, *audio/mpeg* is a compressed audio MPEG file such as an MP3 file and *model/vrml* is an interactive 3D multimedia model file written in Virtual Reality Modelling Language [6, 7, 9].

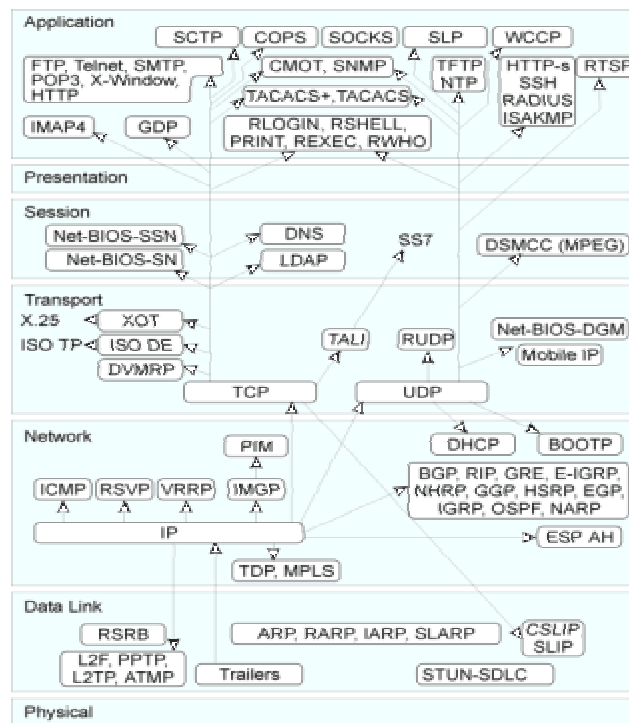


Figure 1. Illustration of Application level protocols (including HTTP) over Transport, Network and other layers (after Protocols.com [5]).

Hypertext Markup Language (HTML).

HTML was developed in 1990 in an attempt to create a cross-platform document publishing language. Its aim was to allow online publishing of documents with particular reference to the layout of these documents so that they might include tables, images and other multimedia, cross-referenced with

hyperlinks. Increasing versatility required from the Web's expansion saw HTML enriched in 1993 with HTML+ and in 1995 with HTML versions 2.0 and 3.0. Further codified versions ensued in 1997 (HTML 3.2) and versions 4.0 in 1998 and 4.01 in 1999. HTML is a data formatting language which allows text, Hypertext and multimedia objects to be explicitly described and supported on multiple platforms. This results in excellent portability across networks and pervasive access to many domains and varieties of information. It is an application of the Standard Generalised Markup Language (SGML) and is therefore a physically organised sequence of characters, ordered logically as a hierarchy of elements [10, 11].

Present to future.

Extensible Markup Language (XML).

With the Web's dramatic growth in recent years have come two main problems: the information can take a frustratingly long time to download due to network congestion, and a specific piece of information can be difficult to pinpoint. These problems reflect shortfalls in the design of HTML in that initially it was only designed as a means to describe the content layout of documents. If content could somehow be endowed with *meaning* then pressure could be removed from the Network, as less supplementary information would need to be exchanged to supply a similar amount of knowledge. This is the creative idea behind XML (and a set of related applications), a new meta-language in development since 1997 designed to supersede HTML and access new levels in the Web's capability [12]. The inflexibility of HTML has led to it becoming over-extended with incompatible additions. XML attempts to redress this by offering non-fixed format use-specific customised mark-up language design, in order to structure data more effectively. Like HTML, the language superclass of XML is SGML, though XML allows more functionality than HTML and is less complex than SGML. Like HTML, XML uses angled bracket tags in pairs called elements to surround the text they describe, although XML tag attributes are not fixed allowing them to be interpreted by the application which is reading them. Elements can be nested to multiple levels to produce a tree structure and thus inheritable and inferential information can be assigned to data structured in this way [13]. This approach endows the meta-language with a marked simplicity and could also lead to useful extended hyperlinks in future WebPages (figure 2). Internationalisation is also supported through the adoption of Unicode standard numerical character assignment. Several co-functional applications designed around XML have been developed in the last few years (figure 3). Scalable Vector Graphics (SVG) is an application designed to extend the multimedia capabilities of the Web. It allows grouping, styling, transformation and importation of vector graphic shapes, text, and images rendered in a dynamic and interactive manner [14]. The Synchronised Multimedia Integrated Language (SMIL) allows users to describe the temporal behaviour

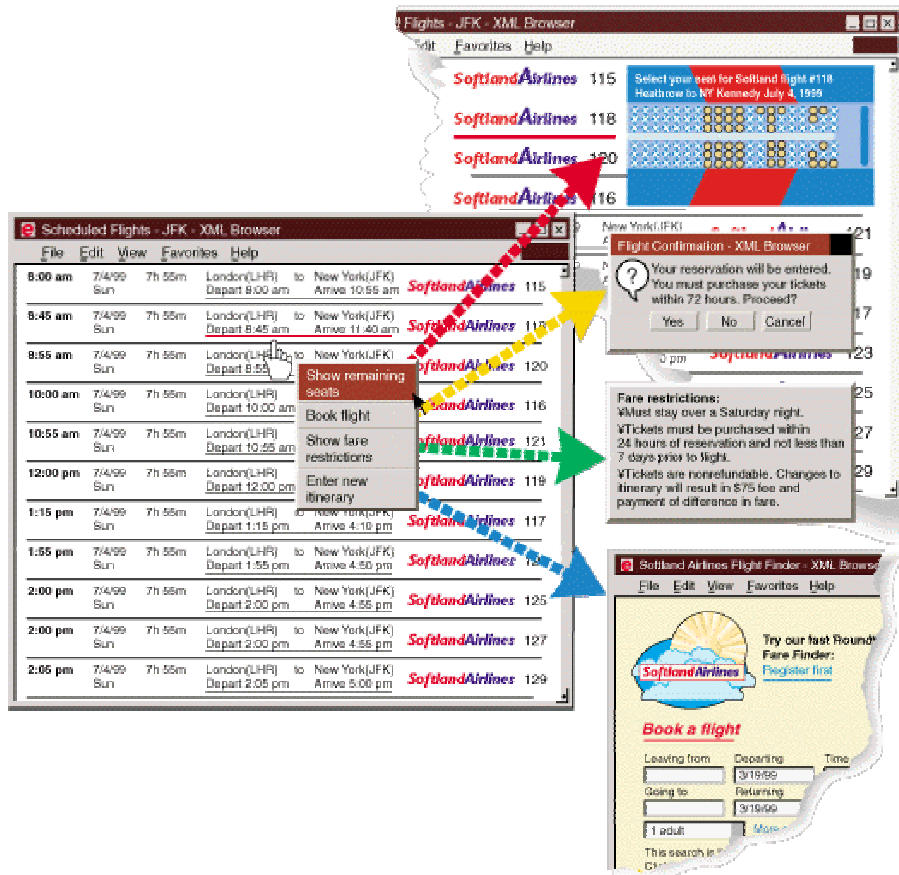


Figure 2. Illustration of how extended hyperlinks could offer a menu of options for more informative navigation with XML (after Bosak & Bray [12]).

and layout of a multimedia representation and to associate hyperlinks with those objects [15]. MathML adds integrated mathematics to the Web framework [16]. XHTML is a reformulation of HTML in XML which attempts to combine, copy and extend HTML with the versatility of XML [17]. XML creates structure in data. The meaning alluded to in the last section can be supplied by applying Resource Description Framework encoding (figures 3 & 4).

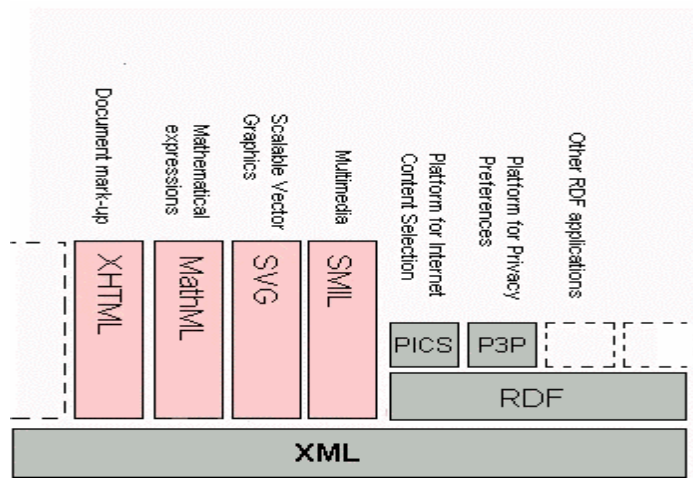


Figure 3. Relation of XML and other developmental applications (after Connolly [18]).

Resource Description Framework (RDF).

RDF has been in development since 1998 and is an application of XML which allows meta-data to be exchanged in a manner which follows a syntactic and schematic standard. Meaning is encoded into meta-data by assigning unique sets of properties to a resource URI within an XML namespace, so that each property has a type and value. These sets of properties are analogous to the subject, verb, and object of a primitive sentence so that resources are linked with properties and other resources. Any resource can thus be described so long as it is identifiable with a URI [19]. Figure 4 shows this relationship. RDF encoding and XML structure point the way to a very enticing possibility in the near future, the Semantic Web.

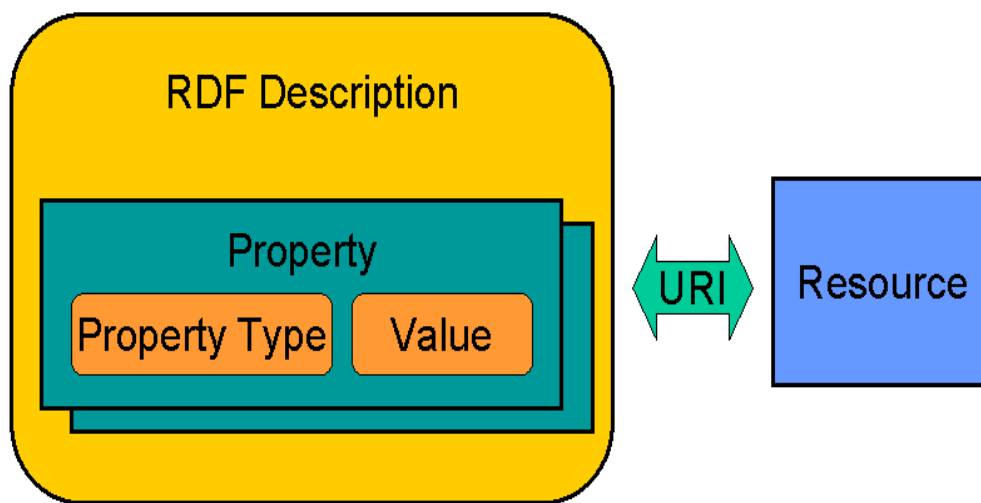


Figure 4. Illustration of relationship between properties and resources in RDF (after Iannella [19]).

Semantic Web.

The Semantic Web idea is the integration of applications such as XML and RDF to produce a powerful information grid of meaningful linked data. The idea requires a third factor to reinforce connectivity, that of Web ontologies. Web ontologies hold properties in formally defined relationships as documents or files pertaining to particular fields of information. This approach is intended to avoid the undesirable ubiquitous broken link of the present Web (figure 5). Ontologies could even point to other ontologies via equivalence relations to strengthen linkage [20]. Software agents will need to be created to do the fetching, processing and exchanging of information. Once this occurs, a Web of structured and standardised meaningful information may be accessed (figure 5). An important factor in this future direction of the Web is the continued upgrade of the physical capability of the Network. The very high speed Backbone Network Service (vBNS) also known as Internet 2 already operates in the US offering packet switching at a minimum of 155 Mb/s projected to deliver 2.2 Gb/s around this time [21], and

References.

- [1] Feizabadi, S. *History of the World Wide Web*. http://ban.mansfield.ohio-state.edu/boat/books/www_beyond_basics/chap1/
- [2] Cailliau, R. *A Short History of the Web*. <http://www.w3.org/pub/WWW/History.html>
- [3] Berners-Lee, T.; Fielding, R.; Masinter, L. *Uniform Resource Identifiers (URI): Generic Syntax* RFC 2396 <http://www.faqs.org/rfcs/rfc2396.html>
- [4] Fielding, R. *et al.* Hypertext Transfer Protocol -- HTTP/1.1 RFC2616 <http://www.faqs.org/rfcs/rfc2616.html>
- [5] Protocols.com WebPages <http://www.protocols.com/pbook/tcpip.htm>
- [6] Nilsson, M. *The audio/mpeg Media Type* RFC3003 <http://www.faqs.org/rfcs/rfc3003.html>
- [7] Freed, N. & Borenstein, N. *Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types*. RFC2046 <http://www.faqs.org/rfcs/rfc2046.html>
- [8] Nelson, S. & Parks, C. *The Model Primary Content Type for Multipurpose Internet Mail Extensions* RFC2077 <http://www.faqs.org/rfcs/rfc2077.html>
- [9] Bell, G. *VRML Specification* http://www.vrml.org/fs_specifications.htm
- [10] Connolly, D. & Masinter, L. *The 'text/html' Media Type* RFC2854 <http://www.faqs.org/rfcs/rfc2854.html>
- [11] Raggett, D. *et al* *HTML 4.01 Specification* <http://www.w3.org/TR/html4/>
- [12] Bosak, J. & Bray, T. *XML and the Second Generation Web* **Scientific American online** <http://130.94.24.217/1999/0599issue/0599bosak.html>
- [13] W3C eXtensible Markup Language version 1.0 Specification <http://www.w3.org/TR/2000/REC-xml-20001006#sec-intro>
- [14] Lilley, C. Scalable Vector Graphics Overview <http://www.w3.org/Graphics/SVG/Overview.htm8>
- [15] W3C Standardised Multimedia Integrated Language version 2.0 <http://www.w3.org/TR/smil20/>
- [16] W3C MathML specification <http://www.w3.org/Math>
- [17] W3C XHTML version 1.0 Specification <http://www.w3.org/TR/xhtml1/#html>
- [18] Connolly, D. XML Activity Page <http://www.w3.org/XML/Activity.html>
- [19] Iannella, R. *An Idiot's Guide to the Resource Description Framework* <http://archive.dstc.edu.au/RDU/reports/RDF-Idiot/>
- [20] Berners-Lee, T.; Hendler, J. & Lassila, O. *The Semantic Web* **Scientific American Online** <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>
- [21] Tynan, D. *Internet2: The Once and Future Net* **MIT Technology Review Online** <http://www.techreview.com/web/tynan/tynan071001.asp>
- [22] GEANT Online Brochure <http://www.dante.net/geant/geant-brochure-nov01.html>
- [23] Brebner, G. (1997) *Computers in Communication, Chapter 9: Case Study 1: Accessing the World Wide Web*. (McGraw Hill, England).