



Instituto Superior de Engenharia de Lisboa

6380 Departamento de Engenharia da Electrónica das Telecomunicações e dos Computadores

Infraestruturas Computacionais Distribuídas WEB

Versão 1.2



Índice

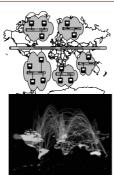
- Introdução Internet/Intranet/Extranet
- Modelo Cliente Servidor
- Serviços
- Serviço WEB
 - Esquema de nomes URI URL
 - Protocolo HTTP
 - Linguagem HTML





Internet - Introdução

- A Internet é uma rede global de comunicações que interliga milhares de Intranets. (token-ring FDDI, Ethernet, etc.)
- Teve origem no exercito dos E.U.A. com a criação da ARPANET em 1969
 - Motivação:
 - Comunicações tácticas na eventualidade dos telefones ficarem inutilizados



ISEL/DEET



Intranet/Extranet

- Intranet Rede interna utilizada pelas organizações para partilharem ficheiros e documentos web (websites).

 - Para uma maior segurança, normalmente não são acessíveis através da Internet.
 Ex A rede do ISEL é interna e não pode ser acedida através da Internet (NAT fornecido pelo proxy).
 Apenas é permitido que os servidores seguin visíveis exteriormente.
- Extranet É uma intranet externa. Ou seja, é uma ligação entre intranets através da internet que através de ligações fiáveis e comunicações protegidas permite redes de negócio para negócio unindo empresas parceiras.
 Por usar os padrões abertos da Internet, os parceiros não precisam usar o mesmo sistema operativo, hardware ou browser.



Arquitectura

Modelo Cliente – Servidor



- Exemplo de Serviços:
 - **■** WEB

ISEL/DEET

- FTP File Transfer Protocol
- eMAIL Electronic Mail

Serviços mais conhecidos

Nome do Serviço	Porto/Protocolo	Descrição
FTP	20,21 / TCP	Transferência de Ficheiros
TELNET	23 / TCP	Shell Remota
SSH	22	Secure Shell, Transferência de ficheiros, Port forward
SMTP	25 / TCP	Envio de Email
POP3	110 / TCP	Consulta de Email
RPC	111 / TCP UDP	Remote Procedure Call
HTTP	80 / TCP	Servidor WEB
HTTPS	443 / TCP	Servidor WEB seguro
NFS	2049 / TCP UDP	Partilha de ficheiros em rede
IRC	6667 / TCP	Internet Relay Chat

1



Serviço WEB - Introdução

- A web é uma aplicação distribuída de grande escala
- Teve origem no CERN no final da década de 80, devido à necessidade dos investigadores partilharem os resultados das suas experiências
- A World Wide Web nasceu quando Tim Berners-Lee, um investigador do CERN, desenvolveu a linguagem HTML (HiperText Markup Language).
- Permite um acesso simples, através de um browser (cliente), a documentos que se encontram em computadores remotos (servidores) ligados à Internet
- Facilita a visualização transparente dos documentos sem a preocupação da heterogeneidade das máquinas que os armazenam

L/DEETC



Serviço WEB - Introdução (cont)

- A World Wide Web (Web) é uma rede de recursos de informação. A Web apoia-se em três mecanismos para tornar estes recursos sempre disponíveis ao maior número possível de pessoas:
 - Um esquema de nomes uniforme para localizar os recursos na Web (ex. URIs, URLs)
 - O protocolo para aceder aos recursos através da Web (ex. HTTP -HyperText Transfer Protocol)
 - Hypertexto, para uma fácil navegação entre os recursos (ex. HTML – HyperText Markup Language)

8



Arquitectura WEB

- Baseado no modelo Cliente Servidor
- A comunicação ao nível de transporte é feita por TCP
- A comunicação obedece ao protocolo HTTP (HyperText Transfer Protocol)
 Pedido/envio de informação ao/para servidor



ISEL/DEETC 9



Interacção Cliente — Servidor

- Interacção Cliente Servidor:
 - O browser interpreta o URL e obtém do DNS o respectivo endereço IP
 - O browser estabelece uma ligação TCP com o servidor Web
 - O browser através do protocolo HTTP envia um pedido do documento em causa
 - O servidor responde enviando o documento pedido e termina a ligação
 - O browser apresenta o documento pedido
- Interacção Cliente Servidor:
 - O servidor Web, por omissão, espera ligações TCP no porto 80
 - Estabelecimento da ligação
 - Servidor recebe o pedido do Cliente
 - Após encontrar o recurso pedido envia-o para o cliente
 - Termina a ligação

EETC



Características dos Servidores Web

- Por omissão espera pedidos TCP no porto 80
 Tem de ser especificado no URL caso contrário
 Ex: http://localhost:8080
- O servidor Web não guarda informação de estado (mensagens idempotentes)
 - Simplifica a sua implementação e reduz os recursos necessários para correr
 - Se houver um problema no servidor, este pode simplesmente reiniciar-se sem precisar de mecanismos elaborados de recuperação
 - Se a comunicação com os clientes deixar de ser possível, o servidor não é afectado
 - Têm de ser utilizados outros mecanismos para manter o estado, dificultando o desenvolvimento das aplicações (ex: comércio electrónico)



Identificadores de Recursos

URI (Uniform Resource Identifiers)

ISLIGETC 11 ISLIGETC



URI

- URI Uniform Resource Identifiers (RFC 2396)
 - Definem a sintaxe para identificadores/localizadores de recursos
- Formato genérico de um URI
 - </pre



- Tipos de URIs:
 - URL Uniform Resource Locators (RFCs 2368, 1738, 1808)
 - @ Identificam por localização
 - URN Uniform Resource Names (RFC 2141)
 - @ Identificadores persistentes, independentes da localização.

ISEL/DEET



URL - Uniform Resource Locator



- Definido pelo RFC1738 disponível em: http://www.ietf.org/rfc/rfc1738.txt
- Caracteres utilizados nos URLs:
 - Caracteres US-ASCII
 - Caracteres que podem ser reservados:
 - "?", ":", "@", "=", "&", "#"]
 - @ Se aparecerem fora do contexto para o qual foram reservados, devem ser codificados
 - Codificação de caracteres: %<código em hexadecimal>
 - Se o esquema tiver links relativos e uma estrutura hierárquica, os componentes da hierarquia devem ser separados por "





URL – Uniform Resource Locator (cont)

- No caso do esquema HTTP o URL tem o seguinte
 - http_URL = "http://"<host>[":" <port>] [<path>] [arguments]
 - ❷ host Nome ou endereço IP de um computador
 - **⊘** port − número inteiro identificador do porto (80/TCP se omitido)
 - @path caminho identificador do documento
 - **@** arguments = "?" <arg>["&" <arg>]
 - @arg = <argName> "=" <argValue>

Ex:

- @ http://www.deetc.isel.pt/
- @ http://phoenix/~dremedios/cgi-bin /hello.cgi?Nome=Diogo+Rem%E9dios

ISEL/DEET



URL – Uniform Resource Locator (cont)

- Para o esquema do FTP temos o seguinte formato:
 - m ftp_URL = "ftp://"[<user>[":"<password>"]@"]<host>[:<port>][/<url-path>]
 @ user UserName do um utilizador
 - user UserName do um utilizador password password bu utilizador referido
 host Nome ou endereço IP de um computador
 port porto a efectuar a ligação (21/TCP se omitido)
 url-path [cwed1>
 url-path [cwed1>
 cwed>/_cwed>/||came>||type=<typecode>|
 cwdx argumento a formeer ao comando cwd (change working directory)
 name nome do ficheiro
 type código com, por exemplo, o tipo de transferência a efectuar texto/binário
 - Ex: ftp://dremedios@phoenix/public_html
- No caso do MAILTO o URL tem o seguinte formato:
 - mailtoURL = "mailto;"[<rfc-822-addr-spec>] [arguments]
 @ arguments = "?" <arg> ["&" < arg >]
 @ arg = <argName >"="<argValue >
 - Ex: mailto:dremedios@deetc.isel.ipl.pt
 Ex: mailto:dremedios@deetc.isel.inl.r
- Para o esquema do emule:
- Ex: ed2k://fileldvdshrink314setup.zip/814771/6CFCA7DC2E0D1706EE34B22CFF44C4DB

ISEL/DEETC



Protocolo HTTP

(HyperText Transfer Protocol)



$Protocolo-HTTP~{\tiny (HyperText~Transfer~Protocol)}$





ISEL/DEET



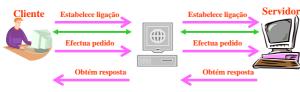
$Protocolo - HTTP \ {\tiny (HyperText\ Transfer\ Protocol)}$

- Protocolo ao nível aplicacional para arquitecturas cliente/servidor
- Protocolo stateless
- É o protocolo usado na Web
- Pensado para permitir o acesso a documentação em formato electrónico.





Funcionamento do Protocolo HTTP



ISEL/DEET



Funcionamento do Protocolo HTTP







Funcionamento do Protocolo HTTP

Proxy:

Pode ser visto como um agente despachante, que recebe os pedidos para um dado URI e rescreve parte ou totalidade da mensagem e reencaminha o pedido para o servidor identificado no URI.





Funcionamento do Protocolo HTTP

■*Gateway*:

É um agente intermediário que permite transferir informação entre servidores que "residem" em camadas diferentes. Se necessário convertem os pedidos para o protocolo do servidor destino.



ISEL/DEETC



Funcionamento do Protocolo HTTP

Tunnel:

Funciona como um ponto de ligação entre dois interlocutores sem modificar o conteúdo das mensagens. É muito utilizado quando o cliente e o servidor estão separados por um firewall.



Funcionamento do Protocolo HTTP

- O pedido do cliente para o servidor e a respectiva resposta é feita em
- As mensagens podem obedecer ao formato MIME (Multipurpose Internet Mail Extensions: RFC 2045), bastando indicar qual a sua versão
- Têm três componentes:
 - **②** Comando (ex: <u>GET</u>, <u>HEAD</u>, <u>POST</u>, PUT, DELETE, TRACE, CONNECT)
 - @ Cabeçalhos:
 - @ Cliente Servidor (indicar a versão do browser, os formatos de documentos que reconhece, etc...)
 - Servidor Cliente (indicar a versão do servidor, dimensão dos documentos, data da última modificação, etc...)
 - Corpo da mensagem



Conteúdo das mensagens HTTP

- Pedido enviado do cliente para o servidor
 - A primeira linha é constituída por:
 - @ Identificação do pedido, denominado o ando ou método
 - Q A identificação do documento em causa (URI)
 - @ A versão do protocolo HTTP
 - As linhas seguintes dependem do método indicado na primeira linha e contêm os cabeçalhos de acordo com o RFC 822
 - III A última linha está em branco (CRLF) indicando o fim do pedido
- Resposta enviada do servidor para o cliente:
 - A primeira linha indica o estado do pedido efectuado
 - Q Se foi servido com sucesso ou , caso contrário, o código de erro indicando o problema
 - As seguintes linhas contêm cabecalhos
 - De seguida virá a informação correspondente ao pedido efectuado (ex: documento)

ISEL/DEETC



HTTP - augmented BNF

<start-line> *(<message-header> CRLF) CRLF [< message-body>]

start-line = <Request-Line> | <Status-Line>

= <Method> SP <Request-URI> SP <HTTP-Version> CRLF Request-Line

= "OPTIONS" | "**GET**" | "**HEAD**" | "**POST**" | "PUT" | "DELETE" | "TRACE" | "CONNECT" | <extension-method> Method

Request-URI = "*" | <absoluteURI> | <abs_path>

Message-header = header-name": "header-value CRLF

Response = <Status-Line> *(general-header | response-header entity-header) CRLF [message-body]

Status-Line = <HTTP-Version> SP <Status-Code> SP <Reason-Phrase> CRLF



Códigos de resposta do servidor

- 1XX Informativo
- 2XX códigos de sucesso 3XX - recolocação/redireccionamento
- 4XX Erro no cliente
- 5XX Erro no servidor

Códigos de resposta

Exemplos

ISEL/DEET

- 200 OK 201 Criado
- 204 Sem conteúdo
- 304 Não modificado
- 400 Pedido mal formado
- 401 Não tem autorização 404 Não encontrado
- 501 Comando não implementado

503 Serviço não disponível

Cabeçalhos HTTP (RFC 822)

- Permite a troca de informação entre o cliente e o servidor
- - Geral Informação que não está directamente relacionada com o cliente, o servidor e o HTTP
 - Pedido Permite indicar preferências nos formatos recebidos e parâmetros do servidor
 - Resposta Informação do servidor que responde
 - Dados Informações referentes aos dados transferidos
- Cabeçalhos gerais:
 - Cache-control Indica como efectuar a cache das páginas no cliente (sim, não e tempo de validade)
 - *Upgrade* Indica a versão do protocolo HTTP a usar
 - MIME-version Indica a versão do MIME a utilizar na comunicação
 - Date Formato preferencial para a data

ISEL/DEETO

5



Cabeçalhos HTTP (cont)

- ☑ Cabeçalhos de pedido (cliente ⇒ servidor):
 ☑ Accept Indica os tipos de dados que sabe interpretar (ex: */*, image/jpeg, text/html, application/msword, etc...)
 ☑ Cookie Permite associar um estado ao cliente. Par "nome" = "valor"

 - If-modified-since indicando uma data, permite informar que só queremos receber o documento caso tenha sido alterado desde então
- Cabeçalhos de resposta (servidor ➡ cliente)
 - Set-cookie Fornece um par "nome"="valor" que fica associado ao cliente permitindo a manutenção de estado
 - Server Informa qual o servidor que está a servir o pedido e qual a sua versão
- Cabeçalhos de dados:

 - Content-length Dimensão em bytes do corpo de dados
 Content-type Tipo de dados contidos no corpo de dados (ex: image/jpeg, text/html,
 - etc...)

 Expires Indica o tempo até quando o documento provavelmente não será alterado e poderá ficar em cache
 - Last-modified Data e hora da última modificação ao ficheiro
 Location Indica que o documento está noutro URL



Métodos GET e HEAD

Método GET

Pede um documento ao servidor dado um URL

(ex: ficheiro, CGI, etc...)

A resposta do servidor contém:

- @ Na primeira linha, o estado
- Nas seguintes os cabeçalhos
- @ E em seguida o documento pedido
- Método HEAD
 - Semelhante ao GET, mas a resposta não inclui o documento
 - Utilizado para saber informações sobre o recurso pedido
 - A resposta do servidor contém:
 - Na primeira linha, o estado
 - Nas seguintes os cabeçalhos

 Útil para saber se a versão que o browser tem em cache já foi alterada •Calcular o tempo de download de determinado documento



Exemplo - GET e HEAD (telnet)



Exemplo GET num browser

Aceder à página:

 $\underline{http://phoenix.deetc.isel.ipl.pt/\sim\!dremedios/index.html}$





ISEL/DEET

Exemplo



Accept: */* Accept-Language: pt

Accept-Encoding: gzip, deflate If-Modified-Since: Thu, 6 Nov 2003 20:10:00 GMT

User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; .NET CLR 1.0.3705)

Host: phoenix.deetc.isel.ipl.pt Connection: Keep-Alive

ISEL/DEETC

33



Exemplo





Método POST

- Método POST
 - Envia dados do cliente para o servidor (no corpo da mensagem)
 - Os dados são enviados no servidor a um programa auxiliar, para serem processados (ex: CGI)
 - O programa a processar os dados é especificado no URL

POST /cgi-bin/post-query HTTP/1.1 Host: phoenix

Accept: "1"
User-Agent: Lynx/2.2 libwww/2.14
From: grobe@www.cc.ukans.edu
Content-type: application/x-www-form-urlencoded
Content-length: 150 org=Academic%20Computing%20Services &users=10000

wsers=mosaic ers=MacMosaic%2C%20WinMosaic

Content-Type: text/html

<h1>Query Results</h1>
You submitted the following name/value pairs org = Academic Computing Services users = 10000 </or>



Métodos PUT, DELETE, TRACE e

- Método PUT
 - Permite colocar no servidor Web, um determinado documento (simétrico ao GET)
 - O URL indica o local onde colocar o documento
- Método DELETE
 - Apaga o documento referenciado no URL
- Método TRACE
 - Permite descobrir, caso seja pedido um dado documento, se no seu trajecto passa por proxy's •Os métodos DELETE, TRACE e CONNECT nem sempre são suportados
- Método CONNECT

Permite fazer tunneling nos proxy's que o suportem

Métodos HTTP



Método	HTTP 1.0	HTTP 1.1
GET	√	✓
HEAD	✓	✓
POST	✓	✓
PUT	X	✓
DELETE	X	✓
TRACE	X	√

Fundamentos do HTML (HyperText Markup Language)

ISEL/DEET



HTML (HyperText Markup Language)

- É uma Markup Language e teve origem na linguagem $SGML(Standard\ Generalized\ Markup\ Language)$
- É uma linguagem que permite descrever a estrutura e a forma como os dados devem ser formatados de forma independente da plataforma
- $\acute{\mathbf{E}}$ baseada em etiquetas (tags) HTML que os browsers interpretam para representar os documentos
- Actualmente o standard é o XHTML $1.0~\rm que$ é uma evolução da versão HTML $4.01~\rm publicado em Dezembro de 1999.$
- Links sobre HTML :
 - http://www.w3.org/MarkUp/
 A especificação está disponível em
 - @ http://www.w3.org/TR/REC-html40/

 - http://www.w3schools.com/html/

 Lista de tags HTML: http://www.w3schools.com/html/html reference.asp
 - http://www.htmlgoodies.com/

HTML (cont)

- Cada etiqueta ou elemento HTML é composto por:
 - ■Uma etiqueta de inicio e outra de fim
 - @ <Elemento> ... </Elemento>
 - ■Pode conter vários atributos

(ç)

Caracteres e

(ã)

- ⟨ELEMENTO atributo1="valor1" atributo2="valor2"> ... </ELEMENTO>
- Os browsers truncam espaços e linhas em branco, contando apenas como um espaço

(ç)

SPACE	·•··) (>)(<) (ė)
		\sim	
Caracteres especiais: &nbs	p;, >,	<,	é,
ã, ç, ç			

ISEL/DEETC



HTML - Estrutura

- A estrutura base de um documento HTML é a seguinte:
 - ... /html> Etiquetas indicativas do inicio e do fim do documento HTML
 - **②** <title> ... </title> cat
 - @ <body> ... </body> Corpo do documento, onde será colocado o conteúdo a ser

🔘 · 📓 🏿 🐔 🔎 Search 🤸



HTML - Estrutura

- Atributos da etiqueta body
 - ■background especifica a imagem de fundo
 - @Ex: <body background="fundo.gif">
 - ■bgcolor especifica a cor de fundo
 - @Ex: <body bgcolor="green">
 - ■bgproperties configura o fundo da página, como fixo ou deslizante
 - @Ex: <body bgproperties="fixed">



HTML - Etiquetas

Etiqueta	Descrição	
<html></html>	Delimita o inicio e o fim do documento	
<head></head>	Delimita a zona de cabeçalhos	
<title></title>	Delimita o título	
<body></body>	Delimita o corpo do documento	
<h_></h_>	Delimita uma secção de nível n 0-6	
ul>	Delimita uma lista não numerada	
<ob< ob<="" td=""><td>Delimita uma lista ordenada</td></ob<>	Delimita uma lista ordenada	
i>	Delimita um item da lista	
 br />	Mudança de linha	
<hr/>	Insere uma linha horizontal	
,	Delimita um parágrafo	
	Insere a imagem indicada no atributo src	
	Delimita um hyperlink para a página indicada em href	
<div> </div>	Define um grupo de elementos (Divisão)	
 	Define um grupo de elementos (inline)	
	Delimita uma zona de comentários	



HTML - Etiquetas (cont)

Etiqueta	Descrição	
<pre></pre>	Delimita texto cujo formato será respeitado mantendo espaços e mudanças de linha	
 	Delimita texto enfatizado	Estilo lógico
 	Delimita texto carregado	\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\
<code> </code>	Delimita código fonte	ν
<i> </i>	Delimita texto em itálico	Estilo
 	Delimita texto em negrito	٧
<object> </object>	Define um objecto embebido (ex animação flash)	

- Applets
 - <a>applet> Permite carregar e correr uma applet (pedaço de código java). Alguns dos seus atributos são: etributos são:

 Code – Indica o código que vai carregar (bytecodes de java)

 - ❷ width e heigth especifica a área que a applet vai ocupar dentro da página

Ex: <applet code= "myapplet.class" width="250" heigth="150" > </applet>

ISEL/DEETO



HTML - Etiquetas (cont)

- Delimita uma tabela. Pode ter os seguintes atributo.
- Stables Delimita uma aucea. Fosc es os agumana.

 Parder Indica a espessura da margem exterior da tabela
 Cellpadding Espaçamento à volta de cada celula
 Width Largura que a tabela vai ter. Pode ter uma percer

 thead> Delimita a zona do cabeçalho dentro da tabela

- | theda-Delimita a zona do cabeçalho dentro da tabela
 | thody-Delimita corpo da tabela
 | tfoot-Delimita of fin da tabela
 | tro-Delimita uma linha da tabela
 | th-Delimita uma coluna que serí cabeçalho da tabela
 | th-Delimita uma coluna dentro de uma linha (célula)
 | th-Delimita uma coluna dentro de uma linha (célula)
 - O atributo Colspan permite indicar que a célula vai ocupar



- Descrição de regras de estilo que indicam ao browser como apresentar os diversos elementos de um documento HTML
- Vêm substituir os atributos de formatação dos elementos HTML.
 - Ex: <hr width="75%" size="5" align="center"/>
- Forma de separar a apresentação do documento da sua estrutura.
- A maioria das propriedades das style sheets podem ser utilizadas em qualquer elemento HTML

ISEL/DEETC



HTML – Cascading Style Sheets (Noções)

Quando um browser lê uma *style sheet*, vai formatar o documento de acordo com ela. Existem três formas de inserir *style sheet's*:

- Style Sheet Externa Ideal para quando o estilo é para ser aplicado a muitas páginas. Torna-se possível alterar o aspecto de todo um web site alterando apenas um ficheiro. Cada documento tem de se ligar à sryle sheet que está a usar, utilizando a etiqueta
 seta etiqueta é colocada na secção dos cabeçalhos. .<head><link rel="stylesheet" type=" 'text/css" href="meuEstilo.css"> </head>
- Style Sheet Interna Uma style sheet interna deve ser usada quando um único documento tem um estilo único. São definidos na secção dos cabeçalhos utilizando a etiqueta <style>.
- Inline Styles Deve ser utilizado quando um estilo é aplicado a uma única ocorrência de um elemento. Para utilizar inline styles é utilizado o atributo style da etiqueta do elemento respectivo. O atributo style pode conter qualquer propriedade CSS. O exemplo mostra como alterar a cor e a margem esquerda de um parágrafo:

This is a paragraph



Exemplo de uma página em html

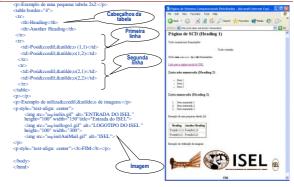


Exemplo de uma página em HTML





Exemplo de uma página em HTML (cont)





HTML - Frames

- A página HTML que contém a definição das molduras não pode ter corpo (<body>)
- Criação de molduras dentro de uma página

 <frameset> Permite definir em quantas regiões vamos querer dividir a janela do browser e a largura de cada uma delas.

 Cob Define a lista com as larguras para as frames verticais
 (ex: -frameset cols = "25%, 25%,*">>)

 Rows Define a lista com as larguras para as frames horizontais
 (ex: -frameset cols = "25%, 25%,*">>)

 Rows Define a lista com as larguras para as frames horizontais
 (ex: -frameset rows = "100, *">)

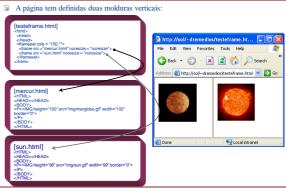
 Border Indica a espessura da linha exterior da muldura ("5" usado por omissão)

 Bordercolor Configura a cor da linha exterior (cinza por omissão)
- <frame> Define a moldura
 - Src Especifica o URL da página que será carregada nesta moldura
 Name Define um nome único para a região
 Frameborder Indica se existirá fronteira na moldura (ex: 0 ou 1)

 - Marginheight, Marginwidth Permite definir em pixeis a margem no Noresize Indica se a moldura pode ser redimensionada pelo utilizador "noresize")

 - Scrolling Indica se a frame tem as barras de scroll (yes, no ou auto)

Exemplo de frames



9



- RFC's
 RFC 1945 HTTP 1.0
 RFC 2616 HTTP 1.1

 - RFC 2045 Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies
 RFC 822 Standard for the format of ARPA Internet text messages

 - RFC 2396 Uniform Resource Identifiers (URI): Generic Syntax
 - RFC 1738 Uniform Resource Locators (URL)

- Special Edition Using XHTML





- HTTP http://www.w3.org/Protocols/
 HTML http://www.w3.org/TR/html401/
 Style Sheets http://www.w3.org/TR/REC-CSS1



HTML^{4.0}