



MIME

■ Multipurpose Internet Mail Extensions

- RFC822 (1982) - formato standard para envio de mensagens de e-mail na Internet
- formato adequado somente para mensagens de texto
- problemas com conjuntos de caracteres mais ricos do que o US-ASCII
- problemas com mensagens multimídia, envolvendo texto formatado, imagens, sons e vídeo
- o padrão MIME foi então definido, de modo que estes problemas pudessem ser contornados, sem entretanto ter que alterar a RFC822
- padrão MIME foi posteriormente utilizado para indicar o conteúdo de páginas Web, semelhantemente a mensagens de e-mail



MIME

■ Normas que descrevem o MIME

- RFC822 - especifica o formato básico para transmissão de mensagens de texto via internet (o MIME deve estar em conformidade com esta especificação)
- RFC2045 - especifica os vários headers utilizados para descrever a estrutura de mensagens MIME
- RFC2046 - define a estrutura geral de sistema de tipos de mídia e define um conjunto inicial de tipos de mídia
- RFC2047 - descreve extensões à RFC822 que permitem o uso de caracteres não-ASCII em headers
- RFC2048 - especifica vários procedimentos de registro para facilidades relacionadas a MIME
- RFC2049 - descreve critérios de conformação à MIME, bem como exemplos ilustrativos de mensagens em formato MIME



MIME

■ Headers MIME

- **MIME-Version** - declara a versão MIME, que ao mesmo tempo informa que a mensagem está no formato MIME e permite que os processadores de e-mail possam distinguir uma mensagem deste tipo de outras mais antigas
- **Content-Type** - utilizado para especificar o tipo de mídia e subtipo de dados utilizado no corpo da mensagem
- **Content-Transfer-Encoding** - especifica as transformações de codificação que foram aplicadas ao corpo da mensagem (normalmente para portá-la para 7 bits ASCII)
- **Content-ID** - usado para identificar o conteúdo do corpo da mensagem
- **Content-Description** - usado para descrever o conteúdo do corpo da mensagem



MIME

■ Tipos de Midia

■ Campo Content-Type define o tipo de mídia

- | tipo/sub-tipo

- | Parâmetros, tais como "charset" podem ser incluídos

■ Tipos básicos

- | text, image, audio, video, application

■ Tipos Compostos

- | multipart,

- Subtipos - mixed, alternative, parallel, digest

- | message

- Subtipos -

- partial - permite a decomposição em múltiplas mensagens

- external-body - permite a referência a um corpo externo à mensagem



HTML

■ Hyper Text Markup Language

- formato não-proprietário baseado em SGML (Standard Generalized Markup Language - ISO8879)
- pode ser criado e processado por uma vasta gama de aplicações

■ Tags

- texto contido dentro dos markups < e >
- <ISTO É UM TAG>
- Não são legíveis
- apresentam instruções de comandos
 - l formatação
 - l acesso a recursos externos (imagens, hyperlinks, código java, etc)



HTML

■ Órgão Regulatório

- W3C - World Wide Web Consortium (www.w3.org)
- MIT Laboratory for Computer Science

■ Versões

- 4.0 (atual - Dezembro de 1997)
- 3.2, 2.0, 1.0
- outras versões não chegaram a se tornar recomendações

■ Features da Versão 4.0

- Suporte ao CSS (Cascade Style Sheets)
- Suporte à Internacionalização
- Suporte aos deficientes físicos
- Tabelas e Formulários mais eficientes
- Suporte à inserção de scripts e multimídia



HTML

■ HTML 4.0 em 3 versões

■ HTML4.0 Transitional

- | utiliza o CSS mas permite também ajustes via markups de formatação de texto - mais flexível

■ HTML 4.0 Strict

- | utilizado para uma implementação puramente estrutural, sem tags associados a layouts, que devem vir em um CSS próprio

■ HTML 4.0 Frameset

- | utilizado quando se deseja dividir a página em diversos frames

■ XHTML 1.0 e HTML 4.01

■ Associa HTML ao XML (Extensible Markup Language)

- Páginas Web mais ricas, podendo ser visualizadas em browsers com diferentes características, tais como WebTV, telefones celulares, kiosks, handhelds, palmtops, etc



HTML

■ Elementos e Atributos

- Elementos são estruturas conceituais hierárquicas encontradas no texto HTML, podendo ser sequenciais e/ou aninhadas
- Elementos são identificados no texto por meio de tags de início e fim. e.g. `<HEAD> ... </HEAD>`
- Alguns tipos de elementos podem não necessitar do tag de finalização. e.g. `<P> ... </P>` -> `<P>`
- Elementos podem conter atributos
 - | ` DCA `

■ Elementos Básicos de um Texto HTML

- HTML, HEAD e BODY

■ Lista Completa dos Elementos no HTML 4.0

- <http://www.w3.org/TR/html40/index/elements.html>



HTML

- **Texto** - Parágrafos, Linhas e Frases
- **Listas** - Listas Não-ordenadas, Ordenadas e Definições
- **Tabelas**
- **Links** - Hipertexto e Links independentes de Midia
- **Objetos, Imagens e Applets**
- **Style Sheets** - Definindo estilos em documentos HTML
- **Alinhamento, fontes e retas horizontais**
- **Frames** - Apresentação com múltiplas janelas
- **Formulários** - Edição de Textos, Botões, Menus, etc
- **Scripts** - Documentos animados e formulários inteligentes



Web Robots

- Web Robots, Spiders, Web Walkers ou Wanderers
 - Agentes de Internet que partindo de uma página web, localizam novas páginas por meio dos elementos **Anchor** inseridos nestas páginas e passam a navegar de página em página.
- Finalidade
 - indexar porções conhecidas da Web
 - localizar links inválidos
 - realizar a manutenção de páginas de um determinado site
 - fazer o cache de páginas potencialmente interessantes
 - filtrar o conteúdo de diferentes mecanismos de busca
 - descoberta de páginas novas e/ou conteúdo novo
 - criação de "mirrors" de páginas com problemas de acesso