INTRODUÇÃO: CONCEITOS BÁSICOS

1.1.1 - WWW

World Wide Web é a "teia de alcance mundial" - "world wide" significa "grande como o mundo" e "Web" significa "teia". A "WWW" foi o mecanismo que permitiu a explosão da Internet porque possibilitou a construção de páginas gráficas, bonitas, coloridas, compostas com as fotos e animações que se podem ver hoje. Além disso, sua maior característica é a possibilidade de navegação através dos "links" (ligação). Pode-se dizer que a WWW é recente, pois surgiu no início dos anos 90 (1992).

1.1.2 - Internet

É a rede das redes. Nasceu após uma experiência militar para conexão de computadores em várias partes do mundo, germinou na experiência de conexão de computadores de diversas universidades e explodiu na conexão de computadores em qualquer lugar do mundo através de uma linha telefônica. A Internet só foi possível depois que inventaram um protocolo de fácil manipulação e que poderia trafegar em qualquer equipamento de informática, o TCP-IP.

A Internet teve seu início nos Estados Unidos em 1969 com o nome de ARPANET, que pertencia ao Depto. de Defesa Norte-Americano.

A partir de 1982 o uso da Arpanet tornou-se maior no âmbito acadêmico, o nome Internet veio depois que a rede se expandiu para outros países (inicialmente países europeus), e chegou ao Brasil em 1989.

Em 1987 foi liberado o uso comercial nos EUA, mas a Internet comercial somente chegou ao Brasil em 1995.

1.1.3 - Protocolo TCP/IP

Transmission Control Protocol/Internet Protocol, ou Protocolo de Controle de Transmissão ou Transferência/Protocolo da Internet, se refere ao conjunto de protocolos utilizados na Internet. Ele inclui uma série de padrões que especificam como os computadores e equipamentos de rede (hubs, switches, roteadores) vão se comunicar e cria convenções para interconectar redes e para o routeamento através dessas conexões.

Características:

- •Estrutura cliente-servidor robusta: O TCP/IP é uma excelente plataforma clienteservidor, especialmente em ambientes WAN (wide-area network, ou redes de grande alcance).
- Compartilhamento de informações: Milhares de organizações militares, educacionais, científicas e comerciais compartilham dados, correio eletrônico (e-mail) e outros serviços na Internet usando o TCP/IP.
- Ampla disponibilidade: Implementações do TCP/IP estão disponíveis em praticamente todos os Sistemas Operacionais populares. Seu código fonte é amplamente disponível em várias implementações.

O protocolo TCP/IP permite que máquinas conectadas numa mesma rede troquem informações entre si, que ocorre na forma de pacotes.

1.1.4 - IP

Existem algumas analogias entre computadores e telefones e o número IP é uma delas. Pode-se imaginar o número IP como um número de telefone com todos os códigos de discagem internacionais. Isto significa que qualquer

máquina pode contactar outra máquina usando o número IP, bastando apenas que exista um caminho entre essas duas máquinas. Para acessar a Internet, toda máquina necessita ter um número IP diferente do número IP das outras máquinas que estejam acessando a Internet naquele mesmo instante. Isto também significa que duas máquinas na mesma rede NÃO podem ter o mesmo número IP, mas duas máquinas podem ter o mesmo IP desde que cada uma

faça parte de uma rede isolada. Essa isolação pode muito bem ser um "firewall" (parede de fogo), que normalmente não impede o acesso dessas máguinas à Internet.

Ainda fazendo analogia com telefones, pode-se imaginar duas pessoas morando em países diferentes que possuam o mesmo número de telefone local, mas que usam prefixos diferentes para as ligações internacionais.

O número de IP tem 4 bytes de tamanho e tem um formato específico, xxx.xxx.xxx (exemplo: 143.107.130.1). Cada agrupamento xxx só pode ir de 0 a 255, pois essa é a capacidade de 1 byte.

1.1.5 - DNS

Domain Name System, ou servidor de nomes, serve para traduzir IPs em nomes de domínios e vice-versa. Por exemplo, quando se quer visitar o "endereço" 143.107.130.1 digita-se FGE.IF.USP.BR, pois é muito mais fácil decorar um nome que um número extenso como um IP, e o servidor DNS traduz esse nome para um IP. Graças ao DNS, os usuários tem o compromisso de saber apenas o nome do domínio que querem visitar e no "header" (cabeçalho) dos pacotes TCP/IP vão os IPs (de destino e de retorno).

No exemplo acima, o usuário digitou FGE.IF.USP.BR e abriu a página do Departamento de Física Geral da USP, isso por que ele enviou pela Internet pacotes TCP/IP que carregavam nos headers o seu próprio IP (para retornarem a ele trazendo a página) e também o IP de destino, 143.107.130.1, que é o local de onde vem página.

Um domínio é composto por um nome e uma extensão, por exemplo USP.BR (IP: 143.107.xxx.xxx).

O DNS é fundamental para o funcionamento da Internet, e é controlado por uma entidade chamada Internic, localizada nos EUA. No Brasil, a autoridade sobre os domínios é a FAPESP (Fundação de Apoio à Pesquisa de São Paulo).

Para ter um domínio na Internet é necessário registrá-lo. Todos os domínios com extensão ".br" são registrados na FAPESP, normalmente através do preenchimento de um formulário de cadastro numa página de "www.registro.br".

1.1.6 - HTTP

Hyper Text Transfer Protocol ou protocolo de transferência de hipertexto é o protocolo utilizado pela WWW. Esse protocolo transmite textos, gráficos e qualquer outro tipo de arquivo além de permitir a navegação através

de hipertexto (HTML). HTTP é o protocolo que tem impulsionado o crescimento rápido e recente da Internet, pois permite o funcionamento da face gráfica, muito mais bonita e amigável, além de sons e animações.

Atualmente o protocolo HTTP é o mais usado do conjunto de protocolos TCP/IP.

1.1.7 - Porta

Servidores HTTP geralmente estão na porta 80, por isto a porta 80 é chamada de porta da Internet, embora o mais correto seria chamá-la de porta da WWW. Exemplos de portas também muito usadas são: 21 (FTP, File

Transfer Protocol), 23 (Telnet, conexão remota), 25 (SMTP, Servidor de email), 53 (DNS), 110 (POP3, para ler emails), 443 (HTTPS, ou HTTP criptografado, seguro).

1.1.8 - Servidor

É o computador equipado com software que permite "servir" a uma rede de computadores. Quanto mais potente o servidor e mais modernos seus programas, maior e melhor poderá ser a rede por ele atendida.

Exemplos de servidores WEB: Apache (Open Source - free, mais usados atualmente), Microsoft Internet Information Server, Netscape Enterprise Server.

1.1.9 - Browser

Para podermos visualizar as páginas (documentos) HTML é usual utilizar um software específico que interprete essa linguagem e apresente os objetos HTML (o conteúdo da página) na tela do computador do cliente (mais precisamente janela browser). É na interpretação do documento HTML que o browser inclui as figuras, mostra cores, links, sons, animações, etc.

Numa definição simples, web browser - ou simplesmente browser - é o Navegador, o software que permite explorar textos, fotos, gráficos, sons e vídeos na Internet. O browser permite pular de uma página para outra com um simples clique nos "links", que são palavras ou expressões grifadas ou qualquer objeto sobre o qual surgir uma mãozinha quando se passa a seta do mouse sobre ela.

O browser foi desenvolvido na Universidade de Illinois (EUA), no começo dos anos 90, e o primeiro chamava-se Mosaic.

Exemplos de navegadores: Netscape, Internet Explorer (mais usados atualmente), Mozilla, Opera, Konqueror.

Um ponto importante a ser notado é que cada browser e cada versão de browser interpreta de maneira um pouco diferente o conteúdo do documento HTML, e que essas diferenças se acentuam à medida que aumenta a complexidade da página.

1.1.10 - HTML

"Hyper Text Mark-up Language", ou linguagem para marcação de hipertexto, é uma linguagem que usa indicações ou marcações no texto ASCII para disponibilizar na Internet imagens, sons ou animações (os documentos HTML são arquivos de texto puro, ou seja, não binários).

HTML é uma linguagem interpretada onde um programa especial (usualmente o browser) interpreta as marcações no texto (tags) e apresenta seus efeitos na tela para realçar, animar, estruturar, formatar e ilustrar o conteúdo do

documento. Uma "tag" é um marcador de formatação de texto.

1.1.11 - Site

Um "lugar", ou seja, um computador ou uma rede de computadores conectada à Internet onde se tem acesso a informações usando um navegador ou outro instrumento de navegação. As páginas da Internet estão nos sites e têm como linguagem básica o HTMI

Numa linguagem simples, um portal é um site com conteúdo tal que atraia muitos clientes HTTP.

1.1.12 - Ferramentas para criar HTML

Como os documentos HTML são arquivos de texto puro, eles podem ser criados com qualquer editor de textos disponível, tais como o Bloco de Notas e o Wordpad (Windows), Edit (DOS), VI, Kedit, Kwrite e Emacs (Linux/Unix), etc. Para que o browser reconheça esses arquivos como documentos HTML eles devem ser salvos com extensões .html ou .htm.

Existem também vários editores HTML WYSIWYG (What You See Is What You Get, o que você vê é o que você consegue) onde a página pode ser criada simplesmente selecionando com o mouse objetos de um menu, e à medida em que é criado o documento já aparece na forma aproximada com que será visualizado pelos navegadores. Porém, neste caso também é recomendável que se conheça um pouco da linguagem HTML antes de usá-los, pois os editores WYSIWYG carregam os documentos com "sujeira". Os editores WYSIWYG são mais apropriados para webdesigners, não para programadores HTML.

Exemplos: FrontPage (Microsoft), DreamWeaver, Netscape Composer (free), Bluefish (free, Linux/Unix).

1.1.13 - HTTP/1.0 e HTTP/1.1

Logo após a criação do protocolo HTTP ele passou a dominar o tráfego de pacotes na Internet. Com o propósito de permitir uma WWW mais rápida e eficiente e remediar as fraquezas do protocolo original

HTTP/1.0 foi criado o protocolo HTTP/1.1 com as principais características:

- A.Suporte a Hospedagem Virtual não IP, ou seja, múltiplas hospedagens de domínios virtuais num mesmo IP (isso é justificado pela carência e custo de IPs).
- B.Como o HTTP pertence ao conjunto de protocolos TCP/IP (pacotes TCP/IP são enviados e recebidos pela rede), o HTTP/1.1 evita algumas fraquezas do TCP/IP e aproveita algumas de suas vantagens. Por exemplo, quando um cliente acessa um documento, uma imagem, etc., HTTP/1.0 cria uma nova conexão TCP que é desfeita após realizada a tarefa, ou seja, fica criando novas conexões o tempo todo, que implica em muito tráfego TCP/IP e pouca informação transmitida. O HTTP/1.1 corrige isso de 2 maneiras:
 - i.Conexão persistente (persistent conexions) permite ao cliente reutilizar a mesma conexão TCP enquanto estiver "conversando" com o mesmo servidor Web;
 - ii.Otimiza a informação transportada pela rede, evitando retornar e buscar outras informações que ficaram para trás. Como resultado temos menos "engarrafamentos de trânsito de pacotes TCP";
- C.Cache mais eficiente: permite buscar a informação que esteja mais próxima ao cliente, sem ter necessidade de retornar novamente a um servidor Web distante;
- D.Autenticação: senhas podem ser enviadas encriptadas, muito mais seguro que simplesmente texto (poderiam haver sniffers no caminho), favorecendo o comércio eletrônico.

1.1.14 - BROWSERS QUE SUPORTAM HTTP/1.1

Apenas as versões 3.0 ou maiores de browsers suportam HTTP/1.1. Por esse motivo, e a julgar pelo rumo que a WWW tem tomado (cada vez mais gráfica, dinâmica, interativa e animada), os browsers versão 2 ou menores parecem mesmo destinados à extincão.

1.1.15 - URL

Uniform Resource Locators especifica a localização dos arquivos nos servidores. A URL inclui o tipo do recurso que está sendo acessado, o endereço do servidor e a localização do arquivo.

A estrutura completa de uma URL é: acesso://host.domain:porta/path/filename, onde

acesso é usualmente do tipo

file um arquivo no disco local;

ftp um arquivo num servidor de FTP anônimo;

http um arquivo da WWW

porta pode ser qualquer uma entre 1 e 65536 desde que esteja disponível pelo Sistema Operacional e que tenha um servidor Web instalado nela. Quando a porta for omitida será usada a porta 80 padrões, que é o "default" do browser e do servidor Web

host (caso haja) é o nome (hostname) dado ao computador na intranet que constitui o domínio "domain"

domain é o nome do domínio

path é caminho ou estrutura de diretório ou subdiretório. A barra "/" é invertida em relação ao DOS ("\") pois é assim que funciona a estrutura de diretórios no Unix, onde iniciou a Internet

filename é o nome do documento. Se "filename" for omitido ou não existir, o servidor Web entregará (caso haja) uma página "default"

Exemplo:

http://fge.if.usp.br/Welcome.html

onde

fge é o nome de um computador (host) da rede do Instituto de Física, if.usp.br é o domínio (domain) e Welcome.html é o nome do documento HTML. Neste caso foi omitida a porta, e por isso usada a 80.

Exemplo de documento HTML simples:

```
<HTML>
<HEAD>
<TITLE>Um Exemplo simples de HTML</TITLE>
</HEAD>
<BODY>
<H1>HTML é fácil de aprender</H1>
<P>Bemvindo ao mundo do HTML.
Este é o primeiro parágrafo. </P>
<P>Este é o segundo parágrafo. </P>
</BODY>
</HTML>
```



Este é o segundo parágrafo.

Para transformar o código acima num documento HTML basta apenas selecionar o conteúdo desde <HTML> até </HTML> e copiar para, por exemplo, o bloco de notas do Windows. Depois disso basta salvar com qualquer nome e extensão .html ou .htm (tomando o cuidado de não deixar o bloco de notas dar a sua contribuição, que é adicionar um .txt no final do nome), e então abrir esse documento com um browser.

NOTA:

HTML não é "case sensitive" (caixa alta ou baixa não faz diferença), por exemplo <title> é equivalente a <TITLE> ou

<TiTlE>, porém existem algumas poucas exceções a essa regra.

IMPORTANTE:

Nem todas tags são suportadas por todos Web browsers, e se um browser não suporta uma tag ele simplesmente a ignora. Entretanto, qualquer texto posicionado entre um par de tags desconhecida é mostrado pelo browser.

Inclusive, alguns browsers conseguem mostrar um arquivo de texto mesmo que as tags <HTML>, <HEAD>, <TITLE> e <BODY> não tenham sido incluídas.

DETALHE:

Para ver uma cópia do arquivo que o browser leu para interpretar e mostrar a página, basta selecionar "View Source" (Ver Código Fonte ou equivalente) num dos menus do browser. Um modo mais rápido e fácil de acessar esse menu é clicar com o botão direito do mouse sobre a página.

Observar como é escrito o código fonte das páginas HTML é uma maneira bem simples de aprender a linguagem HTML, inclusive porque uma vez visulizado o código fonte pode-se salvar o arquivo em disco e usá-lo como ponto de partida para construção de um novo documento HTML.

1.DESCRIÇÃO DAS PRINCIPAIS TAGS

<HTML> - Esse elemento transmite ao browser que o arquivo que ele leu contém informação de código HTML. Na verdade, essa tag é usada por uma questão de padronização, pois a extensão .html do arquivo já garante ao browser que se trata de um documento HTML.

<HEAD> - O elemento HEAD identifica a primeira parte do documento HTML. É no HEAD que se deve incluir, entre outras coisas, o título <TITLE> do documento, que será apresentado como título da página na barra de título da janela browser.

<TITLE> - O Título identifica a página num contexto geral, pois o que for definido na tag <TITLE> será mostrado na barra de título da janela do browser e por isso deve ser escolhido um título único, descritivo e relativamente curto. O título também é usado para identificar a página para os motores de busca (robots), tais como HotBot, Infoseek ou Google (google.com). Geralmente o título deve ter no máximo 64 caracteres.

<BODY> - A segunda (e maior) parte do documento HTML é o corpo, que inclui todo o conteúdo da página. Esse conteúdo é mostrado no interior da janela do browser.

Exemplos de atributos para <BODY>:

A tag

<BODY BGCOLOR="#000000" TEXT="#FFFFFF" LINK="#9690CC"

VLINK="#9690CC" ALINK="#00FF00">

torna o corpo do documento com fundo na cor preta (BGCOLOR), texto branco (TEXT) e hiperlinks prata (LINK) que permanecem na mesma cor após o clique (VLINK - Visited Link). No momento do clique o link adquire a cor verde (ALINK - Active Link).

Em lugar de BGCOLOR poderia ser usado BACKGROUNG.

Exemplo 1.2.2.1:

<BODY BACKGROUND="/util/www/img2/stars.jpg"> que inclui como tema de fundo do documento a figura "stars.jpg".

NOTA 1.

Em HTML as cores RGB (Red, Green, Blue) são obtidas pela combinação de 6 caracteres hexadecimais (base 16) tomados em pares RRGGBB, onde cada par está no intervalo de intensidade 00-FF (cada caracter do par está no intervalo de 0 a 15 na base decimal). Por exemplo, 000000 é totalmente preto (intensidade zero para as três cores), FF0000 é vermelho, 00FF00 é verde, 0000FF é azul e FFFFFF é branco.

Quanto maior o número hexadecimal mais intenso a cor, por isso 0000FF é mais azul que 00000F, pois FF é maior que 0F (na verdade 00000F é quase totalmente preto, pois a contribuição do azul é muito pequena).

Usando esse sistema para as cores é possível representar 256x256x256 cores (16 milhões), pois 16x16=256. Mas convém notar que uma coisa são as cores possíveis de serem apresentadas no documento HTML, outra é a capacidade do monitor do cliente HTTP em reproduzir essas cores.

NOTA 2:

As cores básicas, tais como vermelho, azul, verde, etc. também podem ser representadas pelo seu respectivo nome em inglês, por exemplo <BODY BGCOLOR="red">

TAGS MAIS USADAS NO INTERIOR DO CORPO DO DOCUMENTO HTML

Headings

<Hx>, onde x vai de 1 a 6 (não confundir HEADING com HEAD). Quanto maior o número x, menor o tamanho do texto.
Exemplo:

<BODY>

<H2>Texto "aumentado" de tamanho</H2>

```
</BODY>
```

Paragraphs

<P>, delimita parágrafos no documento.

Vai mostrar na janela browser:

Este parágrafo está centralizado!

IMPORTANTE:

Ao contrário da maioria dos processadores e editores de texto, em HTML espaços em branco, "carriage return" ("Enter") e tabulações são mostrados comprimidos em um único espaço em branco. A vantagem é que não é necessário se preocupar com a extensão das linhas, pois o browser é que vai se encarregar da melhor apresentação do conteúdo do documento na janela. Exceção: tag <PRE> (texto pré-formatado).

NOTA:

Na linguagem HTML toda a formatação do texto é obtida com tags apropriadas. As marcações inseridas pelos editores de texto, tais como quebras de linha, tabulações e parágrafos são desprezadas pelo browser.

Lists

As listas são usadas normalmente para criar índices de páginas HTML, tanto índices numerados quanto não-numerados.

A tag determina listas não numeradas (unnumbered) e determina listas numeradas.

```
Exemplo
       <UL TYPE="circle">
               <LI> maçã
               <LI> pêra
               <LI> laranja
       </UL>
irá mostrar:
       o maçã
       o pêra
       o laranja
Atributos:
<UL TYPE="disc|square|circle"></UL>
<LI TYPE="disc|square|circle">
DISC é um círculo preenchido, SQUARE é um quadrado vazado e CIRCLE é um círculo vazado.
Exemplo
       <0L>
               <LI> morango
               <LI> pêssego
               <LI>uva
       </OL>
irá mostrar
       1. morango
       pêssego
       3. uva
```

```
Atributos:
<OL TYPE="1|a|A|i|I" START="n"></OL>
<LI TYPE="1|a|A|i|I" VALUE="n">
onde,
TYPE="1" é o padrão de numeração, a e A são numerações alfabéticas, i e I são numerações em algarismos romanos e
'n' em START="n" é um número onde se queira iniciar a listagem e VALUE="n" é um valor a ser atribuído à tag <LI>.
Exemplo
        <OL TYPE="1" START="7">
                <LI> banana
                <LI> melancia
                <LI> manga
        </OL>
irá mostrar:
        7. banana
        8. melancia
        9. manga
Exemplo
        <OL TYPE="a">
               <LI> morango
                <LI> abacaxi
                <LI TYPE="I"> pêssego
                <LI>uva
                <LI TYPE="A" VALUE="3"> laranja
        </OL>
irá mostrar:
        a. morango

 b. abacaxi

        III. pêssego
        IV. uva
        C. laranja
```

Nested Lists

Listas aninhadas: consiste em juntar as listas numeradas e não-numeradas para obter um índice mais elaborado.

```
Exemplo 1.2.3.4.1:
       <0L>
        <LI> A few New England states:
               <UL TYPE="circle">
                       <LI> Vermont
                       <LI>New Hampshire
                       <LI> Maine
               </UL>
        <LI> Two Midwestern states:
               <UL TYPE="circle">
                       <LI> Michigan
                       <LI> Indiana
               </UL>
       </OL>
Irá mostrar:
        1. A few New England states:
               o Vermont
               o New Hampshire
               o Maine
       2. Two Midwestern states:
               o Michigan
               o Indiana
```

Definition Lists

Listas de definição: normalmente consistem em alternar um termo de definição, definition term <DT>, e uma definição de definição, definition definition <DD>. Os browsers normalmente formatam a definição numa linha nova com parágrafo (é a tabulação em HTML).

NOTA:

As entradas nas tags <DT> e <DD> podem conter múltiplos parágrafos <P>, listas ou outras informações de definição.

Preformatted Text

Texto Preformatado: a tag <PRE> torna os espaços, linhas e tabulações do editor de textos significantes. É útil para a listagem de programas, entre outras coisas, mas tem a desvantagem de não permitir ao texto "se acomodar" no interior da janela browser caso a linha de texto seja mais extensa que a janela.

```
Exemplo
<PRE>
#!/usr/bin/perl

print "Bemvindo ao mundo do Perl !!!\n";

exit;
</PRE>
irá mostrar

#!/usr/bin/perl

print "Bemvindo ao mundo do Perl !!!\n";

exit;
```

Atributo:

```
<PRE COLS="n" WRAP></PRE>
```

especifica o número máximo de caracteres que serão mostrados na mesma linha. Esse atributo efetivamente quebra ou ajusta (WRAP) as linhas de texto na extensão dada por "n".

OBSERVAÇÃO:

A tag <PRE> não lista as tags HTML, que são interpretadas pelo browser.

Plaintext

Texto Puro: tudo que estiver APÓS a tag <PLAINTEXT> será interpretado como texto puro, incluindo tags HTML que não serão mais interpretadas pelo browser.

```
Exemplo |

<PLAINTEXT>
Exemplo de documento HTML mínimo:
<HTML>

<HEAD>
<TITLE>Documento mínimo</TITLE>
</HEAD>
<BODY BGCOLOR="#FFFFFF">
<P ALIGN="CENTER">Documento HTML mínimo</P>
</BODY>
</HTML>
```

OBSERVAÇÃO 1:

A tag <PLAINTEXT> NÃO tem </PLAINTEXT>, o que significa que após inserida essa tag todas as tags restantes do documento HTML não serão mais interpretadas pelo browser, apenas listadas como texto.

OBSERVAÇÃO 2:

Uma tag similar à <PLAINTEXT> é <XMP>, que tem a vantagem de admitir </XMP>.

Extended Quotations

Citações Estendidas: a tag <BLOCKQUOTE> é usada para incluir texto num bloco separado. Muitos browsers geralmente mudam a margem esquerda para separar a aparência do bloco na janela browser.

```
Exemplo

<P>Omitir palavras desnecessárias:
<BLOCKQUOTE>
<P>"Uma sentença não deve conter palavras desnecessárias, um parágrafo não deve conter sentenças desnecessárias."
<P>--O autor 
</BLOCKQUOTE>
```

irá mostrar

Omitir palavras desnecessárias:

"Uma sentença não deve conter palavras desnecessárias, um parágrafo não deve conter sentenças desnecessárias."

--O autor