# TEXTO, HIPERTEXTO E HIPERMÉDIA

#### **TELEMÉDIA**

#### 2006/2007

### Sumário

- Texto
  - Representação do conteúdo textual
  - Representação da aparência
  - Operações de processamento de texto
- Hipertexto e Hipermédia
  - Âncoras e ligações
  - Um modelo
  - Particularidades do hipermédia em rede

Texto de apoio:

<u>Multimédia e Tecnologias Interactivas</u>, *Nuno Ribeiro*, FCA, 2004 (Secção 4.1 do Capítulo 4)

### Introdução

- Texto é o media mais usado na interacção homem-máquina:
  - Os primeiros monitores dos computadores não exibiam gráficos nem imagens;
  - Os teclados baseiam-se em texto;
- Hoje em dia os elementos gráficos são cada vez mais usados no software, porque:
  - permitem uma interpretação mais rápida
  - ocupam menos espaço no monitor
- Três tipos de texto:
  - Texto n\u00e3o formatado (plain text) texto simples
  - Texto formatado (rich text) inclui a formatação (fontes, etc)
  - Hipertexto (ex: html) texto com ligações a outros textos

#### **Texto**

- Natureza dupla do texto:
  - <u>Conteúdo léxico</u> caracteres que constituem as palavras, sinais de pontuação e outros símbolos;
  - Aparência atributos visuais dos caracteres (fonte, tamanho, disposição no ecran (layout), etc.)
- Identidade fundamental dos caracteres:
  - Caracter abstracto;
  - Representação gráfica;

Por exemplo, o caracter abstracto "A" pode ter uma infinidade de representações gráficos, incluindo " $\mathcal{A}$ ", "A", " $\mathbf{A}$ ", "a", "a", "a"

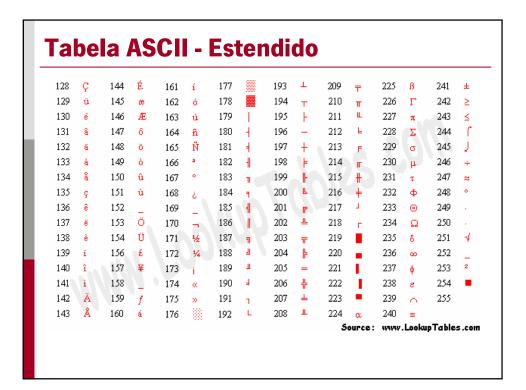
- <u>Caracteres abstractos</u> são os caracteres representados apenas quanto à sua natureza léxica:
  - Caracteres abstractos são agrupados em alfabetos;
  - Cada idioma ou grupo de idiomas usa um alfabeto;
- **Repositórios de caracteres** (*character sets*) tabelas mantidas pelo sistema operativo que consistem numa correspondência entre os códigos e os caracteres
  - Incluem maiúsculas, minúsculas, sinais de pontuação, números e símbolos matemáticos;

### **Texto – caracteres abstractos**

- Vantagens da utilização de <u>repositórios de caracteres</u>:
  - É vital guardar os caracteres na forma de códigos:
    - para poder *editar* (alterar) e *pesquisa* de texto;
    - Para facilitar a *comparação* de caracteres (basta comparar códigos!)
  - Permitem associar os caracteres dos teclados à representação desses caracteres:
    - Por exemplo, quando se pressiona um A no teclado, esse caracter é procurado na tabela de caracteres para depois ser apresentado no monitor
  - Normalização é o mais importante, pois os códigos universais podem facilmente ser trocados entre máquinas distintas e que usam sistemas operativos distintos.

- O primeiro conjunto de caracteres normalizado (1968), foi o **ASCII** (*American Standard Code for Information Interchange*):
  - Utiliza apenas 7 bit... ou seja... 2<sup>7</sup> combinações, portanto um total de 128 caracteres distintos...
  - Chega para o *Inglês*, mas não serve para mais nenhuma língua!...
  - Tentou-se resolver o problema com variantes nacionais a norma ISO 646 – que não foi bem sucedida...
    - As múltiplas variantes contrariam a noção de norma e criam dificuldades na comunicação entre sistemas que usam diferentes variantes da mesma tabela!
  - Solução mais óbvia: usar 8 bit em vez de 7 ASCII Estendido

#### **Tabela ASCII (7bit)** Dec Hx Oct Char Dec Hx Oct Html Chr Dec Hx Oct Html Chr Dec Hx Oct Html Chr 0 000 NUL (null) 1 001 SOH (start of heading) 32 20 040 @#32; Spa 64 40 100 6#64; 96 60 140 4#96; 33 21 041 6#33; 34 22 042 6#34; 2 002 STX (start of text) 66 42 102 &#66: B 98 62 142 6#98; (end of text) (end of transmission) 35 23 043 6#35; # 36 24 044 6#36; \$ 43 103 6#67; 44 104 6#68; 99 63 143 100 64 144 003 6#99: 004 EOT 5 005 ENQ 6 006 ACK (enquiry) 37 25 045 4#37: % 69 45 105 E 70 46 106 F 101 65 145 @#101; (acknowledge) 38 26 046 4#38; 102 66 146 4#102; 103 67 147 6#103; g 104 68 150 6#104; h 007 BEL (bell) 39 27 047 4#39; 71 47 107 6#71; 010 BS (backspace) 40 28 050 @#40; 48 110 6#72; 9 011 TAB (horizontal tab) 41 29 051 6#41; 73 49 111 6#73; 105 69 151 4#105; A 012 LF B 013 VT 42 2A 052 6#42; 43 2B 053 6#43; 74 4A 112 6#74; 75 4B 113 6#75; 106 6A 152 6#106; 107 6B 153 6#107; 10 (NL line feed, new line) 11 (vertical tab) 12 13 C 014 FF D 015 CR (NP form feed, new page) 44 2C 054 , 45 2D 055 -76 4C 114 L 77 4D 115 M 108 6C 154 l 1 109 6D 155 m (carriage return) 14 15 E 016 S0 F 017 SI (shift out) 46 2E 056 . 47 2F 057 / 78 4E 116 N 79 4F 117 O 110 6E 156 n n (shift in) 16 10 020 DLE 17 11 021 DC1 (data link escape) 48 30 060 4#48: 0 80 50 120 6#80; 112 70 160 @#112; p (device control 1) 49 31 061 @#49; 51 121 6#81; 113 71 161 @#113; (device control 2) 50 32 062 4#50; 52 122 4#82; 18 12 022 DC2 82 114 72 162 @#114; 19 13 023 DC (device control 3) 51 33 063 3 3 52 34 064 4 4 53 123 4#83; 115 73 163 s 116 74 164 6#116; 84 54 124 6#84; 20 14 024 DC4 (device control 4) 21 15 025 NAK 22 16 026 SYN 53 35 065 5 5 54 36 066 6 6 85 55 125 6#85; 86 56 126 6#86; 117 75 165 @#117; u 118 76 166 @#118; v (negative acknowledge) (synchronous idle) 23 17 027 ETB 24 18 030 CAN (end of trans. block) 55 37 067 4#55; 87 57 127 4#87: 119 77 167 @#119; W 56 38 070 4#56; 58 130 4#88; 120 78 170 @#120; (cancel) 25 19 031 EM (end of medium) 57 39 071 4#57; 9 89 59 131 4#89; 121 79 171 y Y 58 3A 072 6#58; 26 1A 032 SUB (substitute) 5A 132 @#90; 27 1B 033 ESC (escape) 59 3B 073 4#59; 91 5B 133 6#91; 123 7B 173 { 28 1C 034 FS 29 1D 035 GS 60 3C 074 < < 61 3D 075 = = 92 5C 134 6#92; 93 5D 135 6#93; 124 7C 174 @#124; 125 7D 175 @#125; (group separator) 30 1E 036 RS (record separator) 62 3E 076 >> 94 5E 136 6#94; 126 7E 176 ~ ~ 127 7F 177 DEL 31 1F 037 US 63 3F 077 ? 95 5F 137 @#95; (unit separator) Source: www.LookupTables.com



- O ASCII Estendido n\u00e3o resolveu o problema, porque os 256 caracteres eram insuficientes para todos os idiomas...
- Norma <u>ISO 8859</u> Durante a década de 80 procedeu-se à normalização dos conjuntos de 8 bits:
  - 7 bits (ASCII) + 1 Bit para a variante (ASCII Estendido).
  - Várias variantes regionais (10 variantes);
  - A primeira variante (**ISO 8859-1**), designada por **ISO Latin1**

#### Outras:

- Latin 2 ISO-8859-2 (Europa Oriental: Checo, Eslovaco e Croata)
- ISO-8859-5 (Cirílico); ISO-8859-7 (Grego Moderno) e ISO-8859-8 (Hebreu)

- A opção pelas variantes ISO-8859 acaba por não conseguir resolver bem o problema:
  - 7+1 bit são claramente insuficientes para representar todas as línguas (Chinês e Japonês)
  - E os textos multilingue? Como se trabalha com várias línguas em simultâneo?
- Solução: mais bits!
- Norma <u>ISO-10646</u> (32 bit) de 1991:
  - Permite representar 4.294.967.296 caracteres distintos (2<sup>32</sup>)
  - Desvantagem: qual a diferença entre representar um texto de 50 caracteres em ASCII-Estendido e ISO-10646 em termos de memória ocupada?

#### **Texto – caracteres abstractos**

- Em simultâneo com a norma ISO-10646, um consórcio (Apple, Adobe, Microsoft, HP, IBM, Oracle, SAP, SUN e Unysys) introduziu um conjunto de caracteres de 16 bits:
  - UNICODE
  - 16 bits permitem a representação de caracteres de vários idiomas, incluindo os orientais, no mesmo texto
  - Apesar de permitir representar 65.536 caracteres (2<sup>16</sup>), o UNICODE apenas usa 39.000, deixando os restantes reservados para utilização futura;
    - É compatível com ISO-10646 porque é um subconjunto dessa norma.

#### Conclusão:

Os caracteres abstractos são representados usando repositórios de caracteres, recorrendo a várias normas que têm vindo a evoluir no sentido de ser permitida a escrita de textos multi-língua.



- Uma página WEB ou uma mensagem de correio electrónico podem ser criadas e posteriormente visualizadas em sistemas muito distintos! As representações dos caracteres ganham aí enorme importância!
- Visite a página da disciplina e experimente:
  - Observar no menu "Ver" os códigos escolhidos pelo seu browser;
  - Altere a escolha para Cirilico, Grego e Europa Oriental. Que se passou? Como?
  - Verificar no código fonte HTML ("Ver-> Ver Código Fonte"), logo nas primeiras duas linhas qual o alfabeto (character set) em que foi criada;
- Repita a experiência para outras páginas (Ex: www.google.com)

### Texto - representação gráfica

- Apresentação do texto: a forma dos caracteres
  - A representação visual de um caracteres designa-se por **Glifo**
  - Um caracter pode ser representado por uma infinidade de glifos;
  - Os **glifos** agrupam-se em colecções que se designam por **Fontes**
  - Os glifos de uma fonte partilham um conjunto comum de características visuais (existe uma semelhança entre os glifos da mesma fonte):
    - São desenvolvidos para que combinem bem entre si, tanto no monitor como no papel;
  - As fontes digitais são versões das fontes tradicionais (algumas do séc XV)
  - As <u>fontes</u> podem ser vistas como <u>tabelas de correspondência</u> entre os caracteres abstractos e a sua representação gráfica

### Texto - representação gráfica

- Poderá surgir a seguinte questão:
  - Se
    - 1. Os caracteres abstractos são codificados usando códigos
    - 2. Depois convertidos em representações gráficas para serem apresentados
  - Porque motivo não são logo codificados na sua representação gráfica digital?

### Texto - representação gráfica

- Caracteres de fontes distintas seriam interpretados como caracteres distintos (não existiria a noção de caracter abstracto)
- Uma vez convertidos para forma gráfica, a operação inversa não é tão simples – implica operações de reconhecimento!
  - Não poderia ser facilmente editado, pesquisado, convertido, etc.
- A representação gráfica é ineficiente (consome muito mais memória que os códigos em abstracto);
  - Não permite tirar partido das repetições do mesmo caracter (compressão)

### Texto - representação gráfica

- Onde existem as <u>Fontes</u> que fornecem os <u>Glifos</u> para representar os textos?
- Há duas hipóteses apenas:
  - Ou são guardadas em ficheiros próprios e <u>instaladas no sistema</u> operativo;
    - São partilhadas por todos os ficheiros e todas as aplicações!
    - Quando s\u00e3o requeridas e n\u00e3o existem tem de ser trocadas por fontes alternativas;
  - Ou são <u>embebidas</u> nos próprios ficheiros de texto;
    - Vantagem importante para o designer de uma aplicação multimédia pois é livre de usar qualquer fonte no seu trabalho;
    - Os ficheiros de texto ocupam muito mais espaço;
    - Não se partilham as fontes entre documentos que usam as mesmas fontes;

### Texto - representação gráfica

Exercício com as fontes da Universidade do Minho:



- Descarregue da página da disciplina os ficheiros <u>ExercicioFontes.pdf</u> e <u>ExercícioFontes.ppt</u>. Visualize-os e compare o resultado.
- O ficheiro utiliza a fonte usada na nova imagem da Universidade do Minho: o NewsGotT
  - Procure na pasta "Painel de Controlo -> Tipos de Letra" o ficheiro da fonte respectiva
  - Descarregue-a da página da disciplina e copie-a para essa pasta;
  - Volte a visualizar os dois ficheiros anteriores observando o resultado;
- Verifica nas opções (Ferramentas->Opções) de um dos programas do Office como embeber as fontes:
  - Crie um texto que use as fontes acima e grave-o com fontes embebidas e sem fontes embebidas; verifique as diferenças de tamanho dos ficheiros;

- Mono-espaçadas versus proporcionais
  - Nas fontes <u>mono-espaçadas</u> cada caracter ocupa o mesmo espaço na horizontal, independentemente da sua forma (como nas máquinas de escrever tradicionais)
    - Exemplo: Courier, Courier NEW
  - Nas fontes <u>proporcionais</u> o espaço ocupado depende da largura da forma; As fontes proporcionais produzem texto mais legível que as mono-espaçadas;
    - Exemplos: Garamond, Helvética e Arial

### Texto - classificação das fontes

- Fontes com serif versus Fontes sem serif
  - Os Serif são traços minúsculos que se acrescentam nas extremidades dos glifos; tiveram a sua origem nas marcas produzidas pelos cinzéis dos Romanos nas inscrições em pedra; Daí que também se designam por fontes Romanas
  - Fontes sem serif:
    - Exemplos: Helvética, Arial e Lucida
  - As fontes com serif
    - Exemplos: Garamond, Times New Roman, Century

- Fontes com forma vertical versus Fontes com forma itálica
  - Nas fontes com forma vertical as linhas dos glifos são perfeitamente verticais
  - Nas fontes de forma itálica
    - As linhas verticais são ligeiramente curvadas para a direita;
    - Algumas são apenas variantes das outras fontes usadas para evidenciar texto;
      - Times New Roman em Itálico
    - Outras tentam imitar a caligrafia
      - Exemplo: Lucida Calligraphy, Lucida Handwriting

# Texto - classificação das fontes

- Fontes <u>pesadas</u> versus Fontes <u>leves</u>
  - Fontes pesadas
    - Traço dos glifos é espesso;
    - Podem ser variantes das outras fontes;
    - Não devem ser usadas para frases longas, apenas para destacar palavras;
      - Exemplo: Arial Black
  - Fontes leves
    - Traço dos glifos é fino;

- Fontes para texto contínuo versus Fontes para texto isolado
  - Designação mais subjectiva;
  - Texto contínuo: corpo de um livro ou de um artigo
    - Fáceis de ler;
    - Discretas, não intrusivas;
      - Exemplo: Arial, Times New Roman
  - Texto isolado:
    - Pretende chamar a atenção, fontes intrusivas;
    - Usado em slogans publicitários;
    - Uso de fontes decorativas ou ornamentais;
      - Exemplo: Impact, Comics Sans

### Texto - classificação das fontes

- Ao conjunto de variações de uma fonte (normal, itálico, bold) dá-se o nome de família de fontes;
- Medição das dimensões da fonte:
  - Pontos (pt) 1 pt aproximadamente = 0,3528mm
    - Exemplo 20pt: Dimensão do corpo do caracter, altura entre o topo do caracter mais alto e o fundo do caracter mais baixo;
  - Picas (pc) 1 pc aproximadamente = 4,2333mm
    - Utiliza-se para medir espaçamento entre linhas;

- Quanto á forma como são armazenadas em ficheiro:
  - Fontes **Outline** versus Fontes **Bitmap** 
    - As fontes outline são guardadas sob a forma de gráficos vectoriais;
    - As fontes bitmap são armazenadas como mapas de bits;
    - As fontes outline são mais utilizadas que as fontes bitmap, porque possuem a mesma forma e aparência qualquer que seja a plataforma;
    - O formato mais comum de fonte outline é o <u>Adobe Type 1</u>, sendo os glifos gerados por pequenos programas escritos em PostScript que descrevem curvas Bézier cúbicas; As <u>TrueType</u> são descritas por curvas quadráticas;

### **Texto**

Devem evitar-se passagens longas de texto nas aplicações multimédia porque como são visualizadas no monitor tendem a tornar-se cansativas

#### **Texto**

- Disposição do conteúdo textual (<u>layout</u>)
  - forma como os caracteres se combinam em palavras, frases, linhas, parágrafos e outras unidades de divisão do texto tais como, Secções e Capítulos, isto é, a forma como o texto se dispõe no ecrã ou no papel
  - os documentos formatados possuem uma estrutura interna constituída por elementos estruturais:
    - Cabeçalhos, Títulos, Capítulos, Parágrafos, Listas, Notas de rodapé, Referências, Índices
- Duas formas para definir o layout:
  - Formatos de descrição de estrutura:
  - Formatos de descrição de páginas;

#### **Texto**

- Formatos de <u>descrição de estrutura</u>.
  - Contêm <u>marcas</u> que correspondem à informação de controlo que é adicionada ao corpo do texto (*body text*)
  - Linguagens de **Markup** (baseadas em etiquetas)
    - O SGML (*Standard Generalized Markup*) é um formato de descrição de estrutura usada descrever vários tipos de documentos, por exemplo o HTML.
    - RTF e DOC usados pelo Microsoft Word
    - Latex
  - Simples editores de texto podem criar documentos (Ex: Latex e HTML):
    - WYWIWYG: What You Want Is What You Get
    - Necessário browser ou equivalente para processar e visualizar!
    - Resultado depende do dispositivo de saída: mas é essa a ideia

#### **Texto**

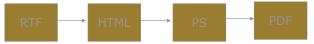
- Processadores de Texto (MS-Word, WordPerfect) ou aplicações desktop publishing (Adobe In Design, Quark Express):
  - WYSIWYG: What You See Is What You Get
  - Todas as operações de formatação são geradas por comandos e visualizadas de imediato!
  - Na verdade, internamente, são geradas e guardadas marcas de formatação!
  - Normalmente só se consegue ver o resultado da formatação com a própria aplicação que a permite gerar;
  - Podem-se exportar para formatos com marcas (ex: HTML)

#### **Texto**

- Formatos de <u>descrição de páginas</u>.
  - São formatos baseados em comandos que descrevem o conteúdo de páginas que podem ser interpretados pelas impressoras ou por uma aplicação que os apresenta no ecrã
  - Exemplos: PS (PostScript) e PDF (Portable Document Format), ambos da Adobe
  - Não é necessário ter software especifico para poder imprimir estes formatos (grande parte das impressoras lê os comandos contidos no ficheiro), embora para poder visualizá-los possa ser necessário um *reader* que não é a mesma aplicação que os produziu



Exercício Conversão de formatos:



• De Rich Text Format para HTML, depois PostScript e finalmente PDF



- Nota: Exercício completo no site
- Produzir um texto RTF (texto.rtf) com o Word; Gravar também em HTML (texto.html) e ver no browser o resultado;
- Instalar uma impressora qualquer PostScript a cores, associada a um ficheiro (ex: HPColorLaser 8500PS que não temos de facto!)
- Imprimir para ficheiro (texto.ps)
- Instalar software "GSView" para ver e converter PostScript
- Ver o PS e converter para PDF
- Identifique os formatos de descrição de página e de estrututura

#### **Texto**

- Operações sobre o texto:
  - Operações sobre caracteres
  - Operações sobre strings (sequências de caracteres)
  - Edição de texto;
  - Formatação de texto;
  - Compressão de texto;
  - Cifragem de texto;
  - Verificação ortográfica e gramatical do texto;

#### **Texto**

#### Conclusão

- O aspecto visual é dado por fontes embebidas no próprio texto ou pré instaladas no sistemas operativo;
- Existem vários tipos de fontes e podem ser classificadas de várias formas;
- •O texto pode ser disposto de várias formas e existem formatos de
- •descrição de estrutura e formatos de descrição de página;
- •Sob o texto podem-se fazer várias operações

# Hipertexto e Hipermédia

Um conceito que engloba:

• a apresentação (forma como o documento é exibido)

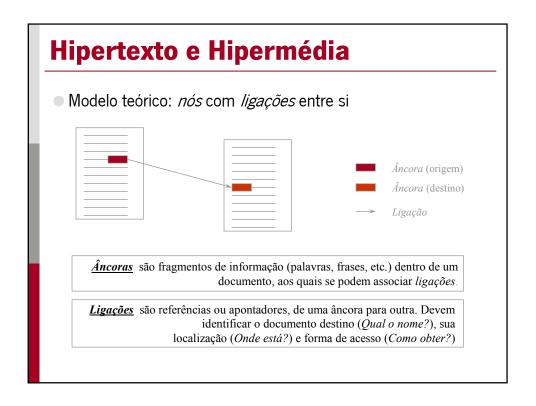
• o acesso (forma de aceder e obter os docs.)

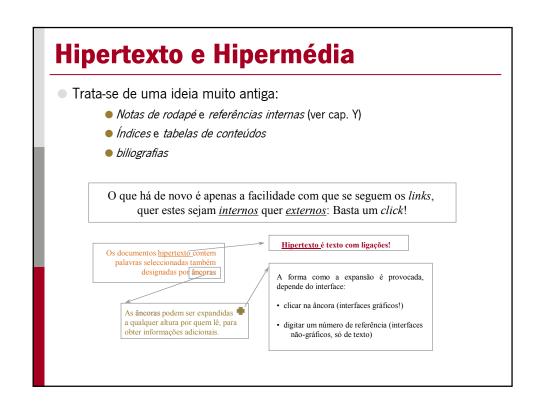
a estrutura (forma com está estruturado)

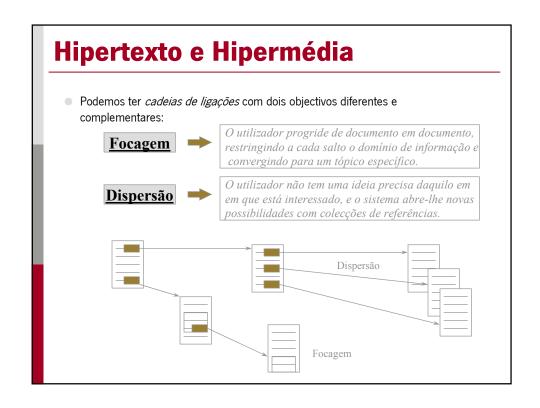
• *o armazenamento* (diferentes formas de armazenamento)

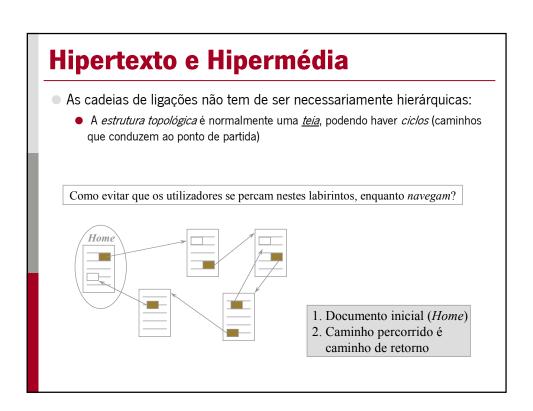
<u>Hipertexto</u> é texto com ligações a outros textos. Os documentos hipertexto não são estritamente sequenciais, pois podem conter *referências* a <u>outras partes do documento</u> ou mesmo a <u>outros documentos</u>.

<u>Hipermédia</u> é média com ligação a outros média. Trata-se da aplicação do conceito de hipertexto, a documentos multimédia.







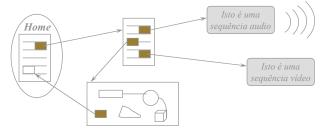


### Hipertexto e Hipermédia

- Outras soluções a que os <u>autores</u> podem recorrer:
  - Mapas com a estrutura, ou diagramas gerais do site
  - Visitas Guiadas previamente preparadas
  - Barras de navegação com sugestões de direcção em cada instante
  - Utilização de <u>Frames</u>, fixando numa parte da janela informação de localização (menus, pág. inicial, etc.)
  - Possibilidade de <u>pesquisa</u> pelo conteúdo em certos pontos da navegação
- Conceitos válidos, independentemente da localização:
  - CD-ROM local, no disco rígido, em servidores distintos, dispersos geograficamente!

# Hipertexto e Hipermédia

Particularidades dos sistemas hipermédia em rede:



- relacionamento <u>temporal</u> entre os objectos!...
- é necessário minimizar tempos de transferência:
  - evitar imagens grandes, sons e vídeo em contínuo...
  - objectos de grandes dimensões só a pedido do utilizador:
    - o notificar os utilizadores do tamanho e do tempo de transferência aproximado!