

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:

Multimédia e Tecnologias Interactivas, 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 não 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:
 - **Conteúdo léxico** – caracteres que constituem as palavras, sinais de pontuação e outros símbolos;
 - **Aparência** – atributos visuais dos caracteres (fonte, tamanho, disposição no ecrã (*layout*), etc.)
- Identidade fundamental dos caracteres:
 - **Caracter abstracto**;
 - **Representação gráfica**;

Por exemplo, o caracter abstracto "A" pode ter uma infinidade de representações gráficas, incluindo "A", "A", "A", "a", "a", "a"

Texto – caracteres abstractos

- **Caracteres abstractos** 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 **repositórios de caracteres**:
 - É vital guardar os caracteres na forma de códigos:
 - para poder *editar* (alterar) e *pesquisar* 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.

Texto – caracteres abstractos

- O primeiro conjunto de caracteres normalizado (1968), foi o **ASCII** (*American Standard Code for Information Interchange*):
 - Utiliza apenas **7 bit**... ou seja... **2⁷** 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	Hex	Oct	Char	Dec	Hex	Oct	Char	Dec	Hex	Oct	Char	Dec	Hex	Oct	Char
0	0000	000	NUL (null)	64	40	060	Space	96	60	140	€#96;	-			
1	0001	001	SOH (start of heading)	65	41	061	!	97	61	141	€#97;	~			
2	0002	002	STX (start of text)	66	42	062	"	98	62	142	€#98;	^			
3	0003	003	ETX (end of text)	67	43	063	#	99	63	143	€#99;	~			
4	0004	004	EOT (end of transmission)	68	44	064	\$	100	64	144	€#100;	~			
5	0005	005	ENQ (enquiry)	69	45	065	%	101	65	145	€#101;	~			
6	0006	006	ACK (acknowledge)	70	46	066	&	102	66	146	€#102;	~			
7	0007	007	BEL (bell)	71	47	067	'	103	67	147	€#103;	~			
8	0010	010	BS (backspace)	72	48	068	(104	68	150	€#104;	~			
9	0011	011	TAB (horizontal tab)	73	49	069)	105	69	151	€#105;	~			
10	A 012	LF	(NL line feed, new line)	74	4A	070	*	106	6A	152	€#106;	~			
11	B 013	VT	(vertical tab)	75	4B	071	+	107	6B	153	€#107;	~			
12	C 014	FF	(NP form feed, new page)	76	4C	072	,	108	6C	154	€#108;	~			
13	D 015	CR	(carriage return)	77	4D	073	-	109	6D	155	€#109;	~			
14	E 016	SO	(shift out)	78	4E	074	.	110	6E	156	€#110;	~			
15	F 017	SI	(shift in)	79	4F	075	/	111	6F	157	€#111;	~			
16	10 020	DLE	(data link escape)	80	50	080	0	112	70	160	€#112;	~			
17	11 021	DC1	(device control 1)	81	51	081	1	113	71	161	€#113;	~			
18	12 022	DC2	(device control 2)	82	52	082	2	114	72	162	€#114;	~			
19	13 023	DC3	(device control 3)	83	53	083	3	115	73	163	€#115;	~			
20	14 024	DC4	(device control 4)	84	54	084	4	116	74	164	€#116;	~			
21	15 025	NAK	(negative acknowledge)	85	55	085	5	117	75	165	€#117;	~			
22	16 026	SYN	(synchronous idle)	86	56	086	6	118	76	166	€#118;	~			
23	17 027	ETB	(end of trans. block)	87	57	087	7	119	77	167	€#119;	~			
24	18 030	CAN	(cancel)	88	58	090	8	120	78	170	€#120;	~			
25	19 031	EM	(end of medium)	89	59	091	9	121	79	171	€#121;	~			
26	1A 032	SUB	(substitute)	90	5A	092	:	122	7A	172	€#122;	~			
27	1B 033	ESC	(escape)	91	5B	093	;	123	7B	173	€#123;	~			
28	1C 034	FS	(file separator)	92	5C	094	<	124	7C	174	€#124;	~			
29	1D 035	GS	(group separator)	93	5D	095	=	125	7D	175	€#125;	~			
30	1E 036	RS	(record separator)	94	5E	096	>	126	7E	176	€#126;	~			
31	1F 037	US	(unit separator)	95	5F	097	?	127	7F	177	€#127;	DEL			

Source: www.LookupTables.com

Tabela ASCII - Estendido

128	Ç	144	É	161	í	177	ÿ	193	±	209	ƒ	225	ß	241	±
129	ú	145	æ	162	ó	178	ÿ	194	ƒ	210	ƒ	226	Γ	242	≥
130	é	146	Æ	163	ú	179		195	ƒ	211	ƒ	227	π	243	≤
131	á	147	ó	164	û	180	†	196	-	212	ƒ	228	Σ	244	ƒ
132	ä	148	ô	165	ÿ	181	‡	197	†	213	ƒ	229	σ	245	ƒ
133	å	149	ò	166	•	182	‡	198	†	214	ƒ	230	μ	246	~
134	â	150	ù	167	•	183	‡	199	†	215	†	231	τ	247	~
135	ç	151	û	168	•	184	‡	200	†	216	†	232	φ	248	~
136	ê	152	-	169	-	185	‡	201	†	217	ƒ	233	Ω	249	~
137	ë	153	Ö	170	-	186	‡	202	†	218	ƒ	234	Ω	250	~
138	è	154	Û	171	¼	187	‡	203	†	219	■	235	δ	251	√
139	í	156	É	172	¼	188	‡	204	†	220	■	236	∞	252	-
140	î	157	ÿ	173	½	189	‡	205	†	221	■	237	φ	253	~
141	ï	158	-	174	»	190	‡	206	†	222	■	238	ε	254	■
142	Ä	159	ƒ	175	»	191	‡	207	†	223	■	239	ε	255	■
143	Å	160	á	176	»	192	‡	208	†	224	∞	240	∞		■

Source: www.LookupTables.com

Texto – caracteres abstractos

- O ASCII Estendido não resolveu o problema, porque os 256 caracteres eram insuficientes para todos os idiomas...
- Norma **ISO 8859** - 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 Latin 1**
- **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)

Texto – caracteres abstractos

- 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 **ISO-10646** (32 bit) de 1991:
 - Permite representar 4.294.967.296 caracteres distintos (2³²)
 - 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 (2¹⁶), o UNICODE apenas usa 39.000, deixando os restantes reservados para utilização futura;
 - É compatível com ISO-10646 porque é um subconjunto dessa norma.

Texto – caracteres abstractos

● 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-lingua.



- 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 Cirílico, 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 **fontes** podem ser vistas como **tabelas de correspondência** 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 **Fontes** que fornecem os **Glifos** para representar os textos?
- Há duas hipóteses apenas:
 - Ou são guardadas em ficheiros próprios e **instaladas no sistema operativo**;
 - São partilhadas por todos os ficheiros e todas as aplicações!
 - Quando são requeridas e não existem tem de ser trocadas por fontes alternativas;
 - Ou são **embebidas** 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 **ExercicioFontes.pdf** e **ExercicioFontes.ppt**. 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;

Texto – classificação das fontes

● **Mono-espaçadas** versus **proporcionais**

- Nas fontes **mono-espaçadas** cada carácter ocupa o mesmo espaço na horizontal, independentemente da sua forma (como nas máquinas de escrever tradicionais)
 - Exemplo: Courier, Courier NEW
- Nas fontes **proporcionais** 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 cinzeiros dos Romanos nas inscrições em pedra; Dai 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

Texto – classificação das fontes

● 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 **pesadas** versus Fontes **leves**

- 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;

Texto – classificação das fontes

● 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 carácter, altura entre o topo do carácter mais alto e o fundo do carácter mais baixo;
 - Picas (pc) – **1 pc** aproximadamente = 4,2333mm
 - Utiliza-se para medir espaçamento entre linhas;

Texto – classificação das fontes

- Quanto à 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 **Adobe Type 1**, sendo os glifos gerados por pequenos programas escritos em PostScript que descrevem curvas Bézier cúbicas; As **TrueType** 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 (**layout**)
 - 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 *descrição de estrutura*:
 - Contêm **marcas** 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):
 - **wywyg**: 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 *descrição de páginas*:
 - 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 específico 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

Texto

- 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 estrutura

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 Hipermedia

- 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)

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

Hipermedia é média com ligação a outros média. Trata-se da aplicação do conceito de hipertexto, a documentos multimédia.

Hipertexto e Hipermedia

- Modelo teórico: nós com *ligações* entre si



Âncoras são fragmentos de informação (palavras, frases, etc.) dentro de um documento, aos quais se podem associar *ligações*

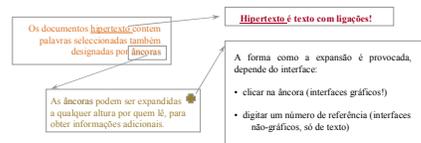
Ligações são referências ou apontadores, de uma âncora para outra. Devem identificar o documento destino (*Qual o nome?*), sua localização (*Onde está?*) e forma de acesso (*Como obter?*)

Hipertexto e Hipermedia

- Trata-se de uma ideia muito antiga:

- Notas de rodapé e referências internas* (ver cap. Y)
- Índices e tabelas de conteúdos*
- bibliografias*

O que há de novo é apenas a facilidade com que se seguem os *links*, quer estes sejam *internos* quer *externos*. Basta um *click!*



Hipertexto e Hipermedia

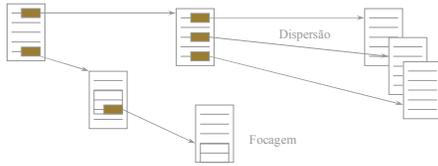
- Podemos ter *cadeias de ligações* com dois objectivos diferentes e complementares:

Focagem

O utilizador progride de documento em documento, restringindo a cada salto o domínio de informação e convergindo para um tópico específico.

Dispersão

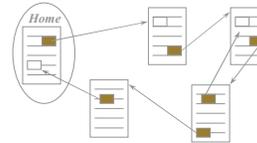
O utilizador não tem uma ideia precisa daquilo em que está interessado, e o sistema abre-lhe novas possibilidades com colecções de referências.



Hipertexto e Hipermedia

- As cadeias de ligações não tem de ser necessariamente hierárquicas:
 - A *estrutura topológica* é normalmente uma *teia*, podendo haver *ciclos* (caminhos que conduzem ao ponto de partida)

Como evitar que os utilizadores se percam nestes labirintos, enquanto *navegam*?



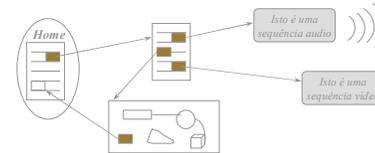
- Documento inicial (*Home*)
- Caminho percorrido é caminho de retorno

Hipertexto e Hipermedia

- Outras soluções a que os *autores* 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 *Frames*, fixando numa parte da janela informação de localização (menus, pág. inicial, etc.)
 - Possibilidade de *pesquisa* 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 Hipermedia

- Particularidades dos sistemas hipermedia em *rede*:



- relacionamento *temporal* 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:
 - notificar os utilizadores do tamanho e do tempo de transferência aproximado!