

Universidade do Minho

Conceitos de Sistemas Informáticos

Módulo 3 – Comunicações por Computador

António Costa <costa@uminho.pt>

Grupo de Comunicações por Computador
Departamento de Informática
Universidade do Minho



Horário de Atendimento:

- Quarta-Feira, 15h – 18h
- Terça-Feira, 10h – 13h

GCOM-DI-UM 1 CSI 2002, A.Costa

Universidade do Minho

Objectivos

- Abordar de forma genérica alguns conceitos de suporte às Redes de Computadores (*Internet*)
 - Forma de funcionamento
 - Aplicações mais utilizadas
- No decorrer de LESI existirão cadeiras onde serão aprofundados conhecimentos mais técnicos sobre a Área das **Comunicações por Computador**
 - *Fundamentos das Telecomunicações (FT)*, *Comunicações de Dados e Redes (CDR)*, *Comunicações por Computador I (CCL)*, *Comunicações por Computador II (CCLII)*, *Sistemas Telemáticos (ST)*

GCOM-DI-UM 2 CSI 2002, A.Costa

Universidade do Minho

AVISOS

- Laboratórios do DI a usar:
 - Terça (11 – 13h) – **DI 1.09**
 - Quarta (14 – 16h) – **DI 0.03**
 - Quarta (17 – 19h) – **DI 0.03**
 - Quinta (16 – 18h) – **DI 1.04**
 - Sexta (14 – 16h) – **DI 1.04**
- Três trabalhos práticos,
 - Como há feriado(s) na quarta-feira, e semana académica, é necessário proceder a alguns ajustes:
 1. Os alunos dos turnos de quarta-feira podem ter de frequentar duas aulas práticas na última semana:
 - o seu turno habitual das quartas-feiras, e,
 - um turno adicional à escolha (terça, quinta ou sexta)
 2. Os alunos dos turnos de terça, quinta e sexta estão dispensados da última semana de aulas práticas, se já tiverem assistido a três aulas, a menos que tenham trabalhos em atraso...

GCOM-DI-UM 3 CSI 2002, A.Costa

Universidade do Minho

AVISOS



GCOM-DI-UM 4 CSI 2002, A.Costa

Universidade do Minho

Conteúdo

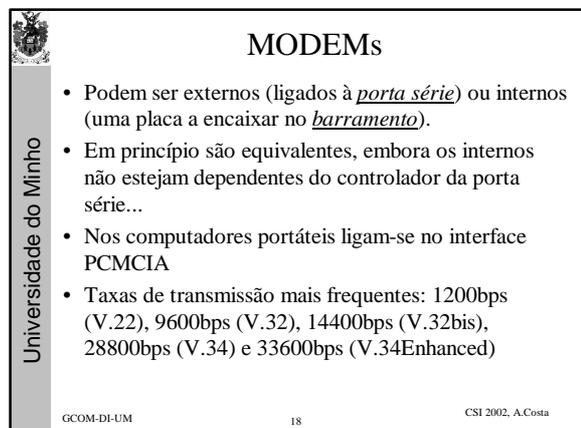
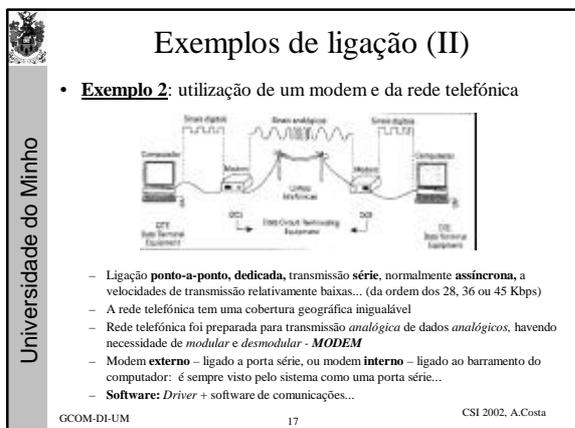
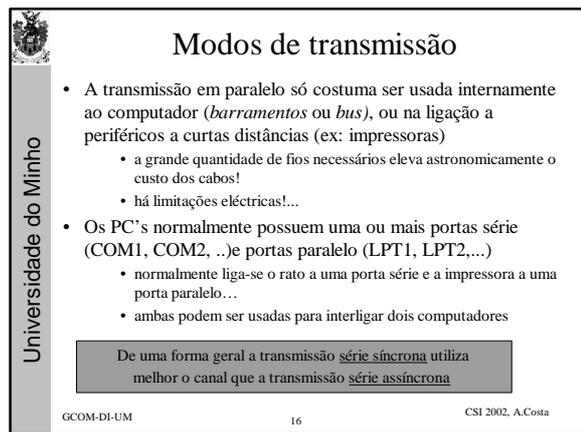
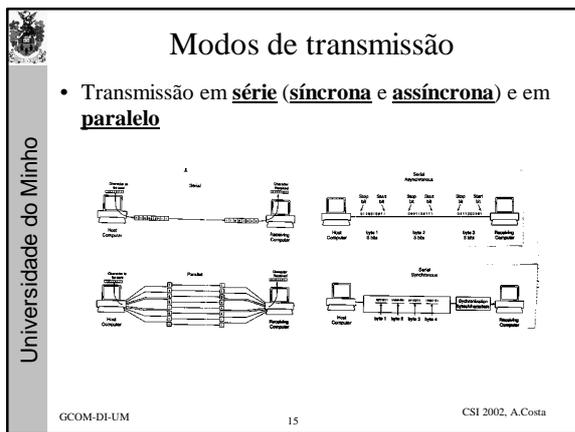
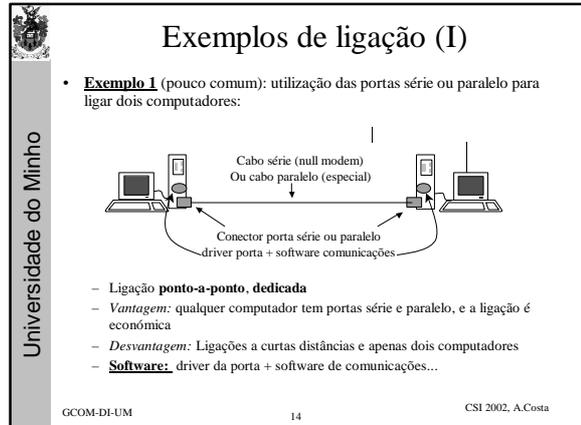
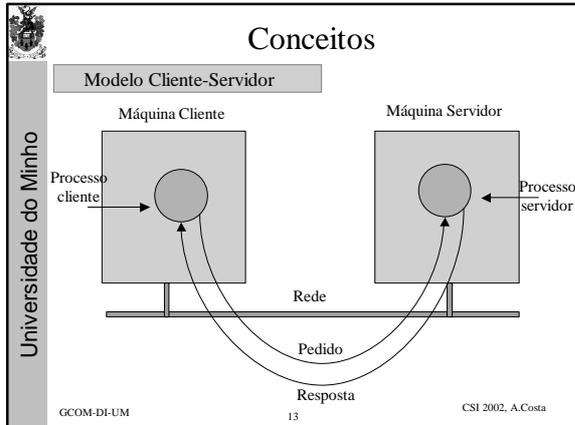
- Introdução às comunicações. Operação da rede através de camadas de protocolos. Noções sobre encapsulamento, endereçamento e nomeação. Definição de protocolo. A pilha protocolar TCP/IP
- Internet. Acesso à Internet. Requisitos e soluções possíveis. Fornecedores de serviço. Protocolos de acesso PPP e SLIP.
- Categorias de aplicações. Correio Electrónico e listas de distribuição. Protocolos IMAP e POP.
- Segurança em Sistemas Telemáticos. Criptografia simétrica e assimétrica. Procedimentos para operações de assinatura digital e confidencialidade. Software PGP.
- Protocolos e linguagens associadas ao WWW: HTTP e HTML. Proxy e caching e a sua importância. Serviços de pesquisa avançada. Segurança e Extensões.
- Comunicações Audio/Video na Internet. Mecanismo de transmissão: tecnologia Multicast. Mbone e Aplicações.

GCOM-DI-UM 5 CSI 2002, A.Costa

Universidade do Minho

Introdução às Comunicações

GCOM-DI-UM 6 CSI 2002, A.Costa



Modos de transmissão

- *Simplex, Half-Duplex, Duplex*

GCOM-DI-UM 19 CSI 2002, A.Costa

Conceitos - Sinais analógicos e digitais

Sinais analógicos

- variação contínua no tempo...

Amplitude

Frequência

Fase

Sinais Digitais

- apenas dois estados:
ON ou OFF, 0 ou 1, etc.

GCOM-DI-UM 20 CSI 2002, A.Costa

Sinais analógicos e digitais

- Técnicas de **codificação**
- e de **modulação**

- Modulação em *amplitude, frequência e fase*

GCOM-DI-UM 21 CSI 2002, A.Costa

Conceitos

Baud Rate (*baud*)

- Número de variações que ocorrem no sinal em cada segundo... Para sinais digitais, 20 Hz correspondem a 20 baud...

Bits por segundo (*bps*)

- Pode ser o mesmo que baud, se uma variação no sinal representar um bit...

Taxa de transmissão

- Medida da quantidade de informação que pode ser enviada por um canal por segundo... (normalmente em bps)

Largura de Banda

- intervalo de frequências de um canal. A taxa de transmissão máxima depende desta diferença entre a freq. máxima e mínima... (ex canal voz: 3.1Kz)

GCOM-DI-UM 22 CSI 2002, A.Costa

Exemplos de ligação (III)

- **Exemplo 3:** ligação a uma rede local do tipo **Ethernet**

- Ligação **multiponto, partilhada**, velocidades de transmissão elevadas... (10 Mbps no exemplo, 100Mbps ou 1Gbps noutros casos)
- Como estão muitos computadores ligados ao mesmo meio físico, é necessário **regras de acesso ao meio físico**, que garantam igualdade de oportunidades para todos!
- Sendo soluções **standard** são também normalmente de baixo custo
- Facilidade de acrescentar e remover sistemas da rede;
- **Hardware:** a placa de rede depende do tipo de rede (**Ethernet, Token Ring, FDDI, etc.**) e do **barramento** interno onde vai ser encaixada (**PCI, ISA, EISA, etc**)
- **Software:** *Driver* da placa de rede (do fabricante) + software de comunicações...

GCOM-DI-UM 23 CSI 2002, A.Costa

Rede local **Ethernet**

- Os dados são transmitidos em pacotes ou *frames*:

Preâmbulo	Endereço destino	Endereço origem	Tipo	Dados	Seqüência de controlo
8 bytes	6 bytes	6 bytes	2 bytes	de 46 a 1500 bytes	4 bytes

- O preâmbulo permite que o receptor se *sincronize* com o emissor
- Cada sistema tem um **endereço** único de 48 bits, atribuído pelo fabricante da placa, que em princípio, não é possível alterar
- Cada *frame* contém o endereço do emissor e do receptor;
- O campo *tipo* é uma espécie de etiqueta que indica que dados são transportados em cada *frame*.
- A *seqüência de controlo* permite detectar erros de transmissão!

GCOM-DI-UM 24 CSI 2002, A.Costa

Universidade do Minho

Redes locais a dois níveis

Lógico	- Pacotes de dados e sua estrutura, endereços, correcção de erros, acesso ao meio físico para recepção e envio...
Físico	- Cabos e tipos de cabos, conectores e suas dimensões, placas de <i>interface</i> , sinais a transmitir, codificação ou modulação dos sinais...

Estes dois níveis podem suportar as aplicações dos utilizadores

Serão suficientes estas duas *camadas*?

GCOM-DI-UM 25 CSI 2002, A.Costa

Universidade do Minho

Exemplos de ligação (IV)

- Exemplo 4:** interligação de redes, utilizando routers

GCOM-DI-UM 26 CSI 2002, A.Costa

Universidade do Minho

Redes alargadas

- Todos podem comunicar com todos, embora não hajam ligações directas entre todos os equipamentos como acontece nas redes locais...
- Há nós intermediários, com *várias ligações*...
 - os equipamentos intermediários designam-se por *routers*...
- Os intermediários necessitam de escolher rotas! (*tomar decisões de encaminhamento*)
- Em caso de falha, é preciso escolher rotas alternativas:
 - só com uma *visão global ou parcial da rede*
- São por isso muito mais complexos e caros!

GCOM-DI-UM 27 CSI 2002, A.Costa

Universidade do Minho

Redes sobre outras Redes

- Um exemplo, com três *camadas*:

GCOM-DI-UM 28 CSI 2002, A.Costa

Universidade do Minho

Redes sobre outras Redes

- Reflexões:
 - A rede de estradas é constituída por *estradas* (de vários tipos) que ligam cruzamentos (ou nós!)
 - À beira das estradas, ou mesmo dos cruzamentos, moram os *utilizadores* nas suas *casas*
 - A rede postal *funciona* sobre a rede viária, e interliga postos que *armazenam e reenviam* correio...
 - A rede postal é uma *rede lógica* sobre a *rede física* das estradas (e pode usar outras redes, como a ferroviária)
 - Não tem de existir um Posto dos Correios em todos os cruzamentos, nem sequer em todas as cidades!
 - Na rede final, no topo, estão os *espíões* (*utilizadores*)
 - Trata-se de uma comunidade com regras próprias, que usa esquemas próprios de *cifragem* da informação
 - Comunicam usando várias redes de suporte: a rede postal, ou mesmo a rede viária directamente...

GCOM-DI-UM 29 CSI 2002, A.Costa

Universidade do Minho

Redes sobre outras Redes

- Três planos:

GCOM-DI-UM 30 CSI 2002, A.Costa

Conceitos

- Isto demonstra alguns conceitos, válidos também nas redes de computadores:
 - a maioria das *redes* funcionam umas sobre as outras

... numa estrutura hierárquica *por camadas*

N
...
1

- uma dada *rede* funciona sobre várias outras, embora seja necessário *interligá-las!*

Interligação de LANs e WANs distintas → Equipamentos:
 * repetidores
 * bridges
 * routers

GCOM-DI-UM 31 CSI 2002, A.Costa

Conceitos

– Interligação de redes

GCOM-DI-UM 32 CSI 2002, A.Costa

Conceitos

– cada *camada* da rede adiciona sucessivamente os seus próprios *conteúdos de informação*:

... o que se designa por *Encapsulamento*

GCOM-DI-UM 33 CSI 2002, A.Costa

Conceitos

– quando os dados mudam para uma rede distinta, de rede para rede, mudam também de *contentor* para *contentor*:

Necessidade de um denominador comum: o pacote IP (*Internet Protocol*)

GCOM-DI-UM 34 CSI 2002, A.Costa

Conceitos

– Cada *camada* usa os seus *endereços* próprios, embora baseados nos endereços das camadas de suporte:

- Os endereços, para além do código postal (dos correios), incluem também o nome da cidade e o nome da rua...
- Numa mesma casa podem morar vários indivíduos, pelo que na morada final se tem de incluir o nome do *utilizador!*

Endereço completo: Av. X + Casa Y + Morador Z

GCOM-DI-UM 35 CSI 2002, A.Costa

Conceitos

• *Endereçamento*

- Exemplo de endereço IP: 193.136.16.254
- Exemplo de nº porta: 80 (porta do servidor WWW!)

Endereço completo (ip+porta): 193.136.16.254:80

GCOM-DI-UM 36 CSI 2002, A.Costa

Universidade do Minho

Arquitectura por camadas

- Modelo de referência OSI (*Open Systems Interconnection*):
 - a comunicação é demasiado complexa para ser monolítica...
 - define 7 (sete!) camadas, independentes:
 - Aplicação, Apresentação, Sessão, Transporte, Rede, Ligação, Física
 - cada camada usa serviços da camada inferior e presta serviços à camada superior
 - Vantagens:
 - cada camada pode evoluir separadamente
 - aplicações mais pequenas e mais rápidas

GCOM-DI-UM 37 CSI 2002, A.Costa

Universidade do Minho

Modelo de referência OSI

Aplicação	Aplicações e serviços distribuídos
Apresentação	Conversão entre formatos utilizados na representação da informação
Sessão	Estabelecimento de conexões entre aplicações
Transporte	Transferência de informação entre dois pontos de forma fiável
Rede	Comunicação fim-a-fim, endereçamento, independência relativamente às camadas inferiores, encaminhamento
Ligação	Regras para a ligação entre dois pontos, detecção e correção de erros, transmissão de blocos de bits
Física	Meio de transmissão, conectores, formato dos sinais, técnicas de modulação, transmissão de cadeias de bits

OSI - Open Systems Interconnection
ISO - International Standards Organization

GCOM-DI-UM 38 CSI 2002, A.Costa

Universidade do Minho

Protocolos

Protocolo:
Conjunto de regras (*sintáticas, semânticas e temporais*) ou convenções que regulam a comunicação entre duas entidades

- Características:
 - directos/índirectos
 - monolíticos/estruturados
 - normalizados (*standard*)/proprietários
 - simétricos/assimétricos

Normalmente surgem agrupados em famílias de protocolos

GCOM-DI-UM 39 CSI 2002, A.Costa

Universidade do Minho

Família de protocolos TCP/IP

Aplicação: FTP, HTTP, SMTP, NFS, DNS
 Transporte: TCP, UDP
 Rede: IP
 Interface ao meio: Rede #1, Rede #2

- No topo de tudo estão as aplicações!
- O protocolo IP (*Internet Protocol*) é o denominador comum...

GCOM-DI-UM 40 CSI 2002, A.Costa

Universidade do Minho

TCP/IP

- O TCP/IP é:
 - “portável”:
 - Funciona em praticamente todos os sistemas operativos
 - Permite endereçamento global (à escala mundial)
 - suportado pela Novell, Microsoft, etc (líderes de mercado)
 - Extensível
 - Totalmente aberto: qualquer vendedor pode escrever a sua própria implementação
- Parte da sua popularidade deve-se ao Unix:
 - Desde o início que foi integrado no Berkeley Unix
 - Usado nas universidades, centros de investigação e agências governamentais (US)

GCOM-DI-UM 41 CSI 2002, A.Costa

Universidade do Minho

IP – Internet Protocol

- Principais funções:
 - Unidade básica para transferência de dados: **pacote IP**
 - Endereçamento: **endereços IP**
 - Encaminhamento: nos **routers** com base no **IP destino** contido em cada pacote; O **IP origem** do pacote não é usado no encaminhamento
- Endereçamento IP
 - Endereço IPv4 - 32 bits (IPv6 –128 bits)
 - xxxxxxxx.xxxxxxxxxx.xxxxxxxxxx.xxxxxxxxxx
 - parte identifica a rede ou subrede, e a outra parte, a interface do *host* com essa rede
 - numa internet, cada endereço tem de ser único
 - é usada notação decimal (Ex: 193.136.9.240)
 - atribuídos pela *Internet Assigned Number Authority*

NOTA: host é qualquer equipamento capaz de transmitir e receber pacotes IP

GCOM-DI-UM 42 CSI 2002, A.Costa

IP – Internet Protocol

Universidade do Minho

Classe A **0** Id de rede (7 bits) Identificação do host (24 bits)

Classe B **1 0** Id de rede (14 bits) Id do host (16 bits)

Classe C **1 1 0** Id de rede (21 bits) Id do host (8 bits)

Classe D **1 1 1 0**

Classe E **1 1 1 1 0**

Classe	Gama de endereços	Número de hosts e de redes
A	0.0.0.0 - 127.255.255.255	126 redes de 16.277.214 hosts
B	128.0.0.0 - 191.255.255.255	16.384 redes de 65.354 hosts
C	192.0.0.0 - 223.255.255.255	2.097.152 redes de 254 hosts
D	224.0.0.0 - 239.255.255.255	Grupos de <i>multicast</i>
E	240.0.0.0 - 247.255.255.255	Reservada para uso futuro

GCOM-DI-UM 43 CSI 2002, A.Costa

Acesso à Internet

Universidade do Minho

GCOM-DI-UM 44 CSI 2002, A.Costa

Acesso à Internet

Universidade do Minho

- É necessário recorrer aos serviços de empresas:

Fornecedor de Serviço Internet
- O acesso implica a escolha:
 - do fornecedor de serviço
 - do tipo de ligação até ao fornecedor de serviço

GCOM-DI-UM 45 CSI 2002, A.Costa

Acesso à Internet

Universidade do Minho

Que tipo de ligação?

- A escolha do *tipo de ligação* depende das características da organização ou empresa:
 - ligações permanentes
 - destinatários: grandes empresas ou universidades
 - maior custo, maior capacidade de transmissão
 - acessibilidade 24h/24h
 - endereços IP alocados permanentemente
 - ligações não-permanentes
 - destinatários: PME e utilizadores individuais
 - menor custo, menor capacidade de transmissão
 - acessível apenas quando "ligado"
 - endereços IP atribuídos dinamicamente

GCOM-DI-UM 46 CSI 2002, A.Costa

Acesso à Internet

Universidade do Minho

Suporte de ligação?

- Rede telefónica (PC + *modem* + linha telefónica)
 - trata-se da solução mais comum...
 - baixas velocidades, baixos custos
- Rede Digital de Integração de Serviços (RDIS)
 - serviço telefónico + transferência de dados
 - linha digital, 2 x 64 *Kbps!*
- Linhas dedicadas ponto-a-ponto
 - locais, inter-urbanas ou internacionais
 - custo em função da distância e da capacidade em *bps*
- Rede da televisão por cabo (TV Cabo, Bragatel)

GCOM-DI-UM 47 CSI 2002, A.Costa

Acesso à Internet

Universidade do Minho

Que fornecedor de Serviço?

- Internet Service Providers* (ISP)
- A escolha do *tipo de fornecedor de serviço de acesso* deve ter em conta:
 - as facilidades disponibilizadas:
 - Número e distribuição geográfica pelo país dos pontos de presença (PoP – Points of Presence)
 - Qualidade das ligações nacionais e internacionais a outros fornecedores de serviço (ISPs nacionais ou internacionais)
 - os custos envolvidos:
 - custos fixos (mensalidades)
 - custos variáveis (dependentes do tempo de ligação ou do volume de tráfego)

GCOM-DI-UM 48 CSI 2002, A.Costa



Resolução de Nomes (DNS)

Universidade do Minho

GCOM-DI-UM 50 CSI 2002, A.Costa

Resolução de Nomes (DNS)

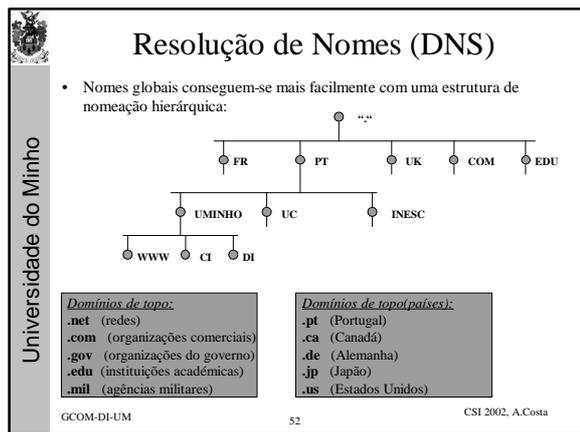
Universidade do Minho

- Nomeação
 - Problema: endereços IP não são fáceis de memorizar!
 - É mais fácil memorizar características como a *côr*, a *marca*, a *localização*, a *tarefa* que desempenha, etc..

Em vez de endereços, podemos baptizar as máquinas com nomes, como *cinzento*, *LaserHP*, *MacCP11*, *WWW*, etc...

- precisamos de uma espécie de lista telefónica com nomes e endereços: DNS (*Domain Name System*)
- Nomes locais* não servem, precisamos de *nomes globais*!

GCOM-DI-UM 51 CSI 2002, A.Costa

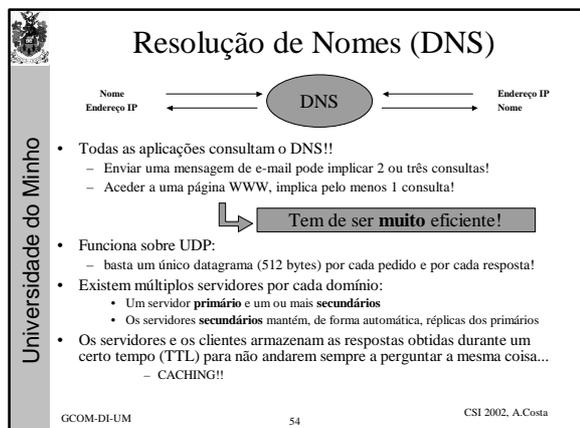


Resolução de Nomes (DNS)

Universidade do Minho

- Nomes são obtidos por concatenação:
 - Exemplos: *www.uminho.pt*, *ci.uminho.pt*
- É impossível saber há partida se se trata de um *nome de sistema*, ou simplesmente de uma *organização* (domínio)
- É necessário uma *Base de Dados* com todos os nomes e endereços do mundo:
 - Necessariamente **distribuída!** (cada qual gere a sua parte)
- O serviço DNS permite consultar a informação, mas nem sequer damos por ele...

GCOM-DI-UM 53 CSI 2002, A.Costa



Universidade do Minho

Resolução de Nomes (DNS)

- Como se “distribui” a base dados?
 - De acordo com a estrutura hierárquica...
 - Um domínio **delega autoridade** aos sub-domínios...
 - Cada domínio é mantido e gerido por uma entidade
- Com tantos servidores, como se fazem as consultas?
 - Cada sistema precisa conhecer apenas um único servidor de DNS (normalmente o do seu domínio!)
 - É a esse servidor que as aplicações cliente formulam as suas *queries*
 - Qualquer servidor conhece sempre outros servidores!
 - Um servidor responde sempre a uma *query*:
 - Ou dá uma resposta.... Ou indica outro servidor que o possa fazer...
 - Um cliente obtém sempre uma resposta válida:
 - Nem que tenha de interrogar – automaticamente – vários servidores!

GCOM-DI-UM 55 CSI 2002, A.Costa

Universidade do Minho

E-Mail (SMTP e MIME)

GCOM-DI-UM 56 CSI 2002, A.Costa

Universidade do Minho

E-Mail

Endereços

- Cada utilizador precisa de ter um *endereço* para poder trocar mensagens... e tem de ser único!
- Os endereços de E-Mail aproveitam os nomes de domínios e de sistemas mantidos no DNS...
- Formato de endereço:


```
<nomeutilizador>@<departamento>.<organização>.<país>
```
- Duas variantes:
 - *user@dominio* ou *user@computador.dominio*
 - A primeira forma é preferível e a mais usada actualmente...
- Exemplos:
 - *costa@uminho.pt*, *macedo@uminho.pt*, *osg@di.uminho.pt*

GCOM-DI-UM 57 CSI 2002, A.Costa

Universidade do Minho

E-Mail

Funcionalidades básicas

- Escrever, Ler, Imprimir e Apagar mensagens
- responder a uma mensagem (*reply*)
 - a mensagem original pode ser parcial ou totalmente incluída, distinguindo-se pela inclusão de um carácter > no início de cada linha
- reenvio (*forward*)
- ordenar o correio em *Pastas (folders)*
- imprimir, gravar, remover mensagens
- agenda de endereços
 - adicionar, pesquisar e remover endereços de e-mail

GCOM-DI-UM 58 CSI 2002, A.Costa

Universidade do Minho

Formato das Mensagens

- Protocolo **MAIL**, definido pelo RFC 822
 - Bastante simples e permite apenas texto
- Cada mensagem tem duas partes:
 - um cabeçalho e um corpo separados por uma linha em branco
- O corpo é texto plano com comprimento limitado
- O cabeçalho é um conjunto de campos estruturados com informação sobre a mensagem
- Cada linha é um campo diferente a menos que comece com um ‘’, que marca uma linha de continuação

GCOM-DI-UM 59 CSI 2002, A.Costa

Universidade do Minho

Formato das Mensagens

- Formato:

From: <endereço do originador>		
To: <endereço do destinatário>		
Cc: <cópia de cortesia>		
Bcc: <cópia de cortesia "cega">		
Subject: <assunto>		
-----texto da mensagem-----		
- Exemplo:


```
To: laura.smith@umich.edu
Cc:
Subject: Citation Verification
-----
Laura, please get me a copy of the following article for which, unfortunately, I only have the author and year of publication: (Jonsey, 1993). As usual, I need the information yesterday.Thanks.
```

GCOM-DI-UM 60 CSI 2002, A.Costa

Universidade do Minho

Envio das mensagens: SMTP

- Protocolo **SMTP** (Simple Mail Transfer Protocol)
- Protocolo extremamente simples para transferir uma mensagem de uma máquina para outra, mas apenas texto:
 - limites no tamanho máximo de cada linha: 1000 caracteres
 - só permite caracteres ASCII de 7 bits ASCII

→ Impossível transmitir binários, ou outros *media* estruturados;

- **Soluções:**
 - Codificação dos conteúdos binários usando codificadores *uuencode/uuencode*
 - Extensões ao protocolo SMTP que ultrapassem as limitações
 - » nem todos os servidores as suportam!
 - Exemplos: transporte a 8bit; negociar tamanho linha; etc.
 - **MIME !!**

GCOM-DI-UM 61 CSI 2002, A.Costa

Universidade do Minho

MIME

- **Multipurpose Internet Mail Extensions:**
 - protocolos que permitem a inclusão de objectos dentro de mensagens, mantendo total compatibilidade com os formatos RFC822 mais antigos
- Modo de funcionamento:
 - definição de 5 novos campos de cabeçalho:
 - MIME-Version, Content-Type, Content-Transfer-Encoding, Content-Description, Content-ID
 - cada mensagem MIME tem sempre definido o seu tipo de conteúdo:
 - Exemplo:
 - » Content-Type: multipart/mixed
 - definição de 5 formas de codificação do conteúdo para envio a 7 bits:
 - 7-bit, 8-bit, binary, quoted-printable, base64
 - Exemplo:
 - » Content-Transfer-Encoding: base64

GCOM-DI-UM 62 CSI 2002, A.Costa

Universidade do Minho

Tipos de Conteúdo

- Exemplos dos principais tipos MIME
 - **Texto:**
 - *text/plain* (texto não estruturado), *text/html* (página html), etc.
 - **Imagem:**
 - *image/gif*, *image/jpeg*, etc.
 - **Audio:**
 - *audio/mp3*, *audio/midi*, etc.
 - **Video:**
 - *video/mpeg*
 - **Aplicação:** formatos manipulados por aplicações
 - *application/postscript*, *application/pdf*, etc.
 - **Multiparte:** incluir várias partes de diferentes tipos
 - *multipart/mixed* (para visualização sequencial)
 - *multipart/parallel* (para visualização em paralelo)
 - *multipart/alternative* (conteúdo repetido em formatos alternativos)

GCOM-DI-UM 63 CSI 2002, A.Costa

Universidade do Minho

Acesso a caixas de correio remotas

POP - Post Office Protocol	IMAP - Internet Message Access Protocol
<ul style="list-style-type: none"> • funções básicas de manipulação de 1 mailbox • processamento local do mail • 3 fases: <ul style="list-style-type: none"> • autenticação (login e password) • acesso à mailbox (LIST, RETR) • quit (actualiza a mailbox) 	<ul style="list-style-type: none"> • permite manipular múltiplas mailboxes • mantém mail no servidor - disponibilizando funções gestão remota • suporta três modelos de mail: <ul style="list-style-type: none"> • <i>offline</i> • <i>online</i> • <i>disconnected</i>

GCOM-DI-UM 64 CSI 2002, A.Costa

Universidade do Minho

Outros: Listas e News

GCOM-DI-UM 65 CSI 2002, A.Costa

Universidade do Minho

Listas distribuição

GCOM-DI-UM 66 CSI 2002, A.Costa

Universidade do Minho

Listas distribuição

- Para que servem?
 - Grupos de discussão
 - Projectos
 - Mailing para clientes
 - Jornais electrónicos
- Tipos de listas de distribuição
 - Inscrição
 - fechadas
 - abertas
 - Envio de mensagens
 - moderadas
 - não moderadas

GCOM-DI-UM 67 CSI 2002, A.Costa

Universidade do Minho

Listas distribuição

- Inscrição numa lista
 - Enviar uma mensagem para
 - *nome-da-lista-request@endereço* com o seguinte conteúdo
 - subscribe <Primeiro Nome> <Último Nome>
 - *listmanager@endereço* com o seguinte conteúdo
 - subscribe *nome-da-lista* <Primeiro Nome> <Último Nome>
- Outros comandos para gestores de listas
 - help [topic]; set <list> [-option] <arg(s)>; unsubscribe <list>; signoff <list>; recipients <list>
 - **ou**
 - information <list>; statistics <list>; run <list> [-password] <cmd [args]->; lists; index [archive | path-to-archive] [password] [-all]

GCOM-DI-UM 68 CSI 2002, A.Costa

Universidade do Minho

USENET NEWS

- Grupos de interesse onde se discute tudo e mais alguma coisa (mais de 5000)
 - milhares de participantes em cada grupo
 - Mensagens similares às de correio electrónico trocadas aos milhares por dia
- Muitas listas de distribuição também são grupos de news
- As News são mais eficientes que as listas de distribuição quando há milhares de utilizadores, milhões de mensagens e a informação não é confidencial...
 - São também por vezes menos úteis por excesso de participantes...
- Com as News não há necessidade de inscrição centralizada
 - Todos os grupos são abertos
 - Tal como nas listas há grupos moderados e não moderados

GCOM-DI-UM 69 CSI 2002, A.Costa

Universidade do Minho

NEWS

- Organização dos Grupos de News
 - Estão organizados hierarquicamente com grupos, subgrupos e sub-subgrupos
 - O nome de cada grupo está separado do seu pai e dos filhos por um (.), por exemplo: *soc.culture.portuguese*
 - Hierarquia das News: primeiro nível
 - *comp* Tópicos relacionados com computadores e informática
 - *news* Rede e software de news
 - *rec* Hobbies, actividades recreativas e artes
 - *sci* Investigação científica e aplicações
 - *soc* Aspectos sociais
 - *talk* Debate em assuntos controversos
 - *misc* Algo que não caiba nos anteriores
 - *pt* Tópicos em português

GCOM-DI-UM 70 CSI 2002, A.Costa

Universidade do Minho

Segurança

GCOM-DI-UM 71 CSI 2002, A.Costa

Universidade do Minho

Segurança

- Computadores podem ser alvo de ataques:
 - ler informação confidencial
 - produzir alterações na informação armazenada
 - destruir o sistema!
- Quando ligados a uma rede:
 - Também a comunicação de dados pode ser objecto de ataques!
 - Já não é necessário acesso físico ao equipamento
 - O número de potenciais atacantes torna-se maior...

GCOM-DI-UM 72 CSI 2002, A.Costa

 **Segurança**

Universidade do Minho

- A introdução de mecanismos de segurança
 - visa minimizar as vulnerabilidades do sistema, mas...
 - ... torna os sistemas mais caros e mais difíceis de usar
- É conveniente:
 - Identificar as vulnerabilidades do sistema
 - Identificar os ataques que podem explorar essas vulnerabilidades
 - Estimar o custo de cada ataque, se concretizado
 - Estimar o custo das contra-medidas a adoptar
 - Fazer uma análise custo/benefício para decidir que mecanismos integrar no sistema

GCOM-DI-UM 73 CSI 2002, A.Costa

 **Ameaças mais comuns**

Universidade do Minho

- Ameaças activas
 - se concretizadas produzem alteração da informação (armazenada ou transmitida)
- Ameaças passivas
 - não produzem alteração de informação, estado ou operação dos computadores
 - obtenção de informação a usar na concretização de ameaças activas

GCOM-DI-UM 74 CSI 2002, A.Costa

 **Ameaças mais comuns**

Universidade do Minho

- disfarce
 - um utilizador faz-se passar por outro...
- interceptação de dados
 - os dados armazenados ou trocados numa comunicação são observados por utilizadores não autorizados...
- interceptação de identidade
 - a identidade de um ou mais utilizadores em comunicação é observada para uso indevido... tipicamente para o atacante se “disfarçar”...
- repúdio
 - Utilizador nega ter realizado determinada acção que de facto realizou... por exemplo ter participado numa comunicação

GCOM-DI-UM 75 CSI 2002, A.Costa

 **Ameaças mais comuns**

Universidade do Minho

- manipulação
 - os dados que fluem na comunicação são manipulados (substituição, inserção, remoção ou alteração da ordem); mesmo informação armazenada, como os programas, podem ser manipulados...
- negação do serviço
 - atrasar ou mesmo impossibilitar a operação da rede ou dos sistemas a ele conectados (por exemplo gerar automaticamente falsas conexões com um servidor!)
- exploração de erros do sistema operativo e do *software*,
- etc.

GCOM-DI-UM 76 CSI 2002, A.Costa

 **Serviços de Segurança**

Universidade do Minho

- Autenticação
 - oferece protecção em relação ao *disfarce* e *intercepção de identidade*
 - exemplos:
 - verificação de identidade na ligação de um utilizador a um computador remoto por *telnet*;
 - verificação de identidade do originador (*From:*) de uma mensagem de correio electrónico
- Controlo de acessos
 - protege o uso não autorizado de recursos disponíveis (leitura, escrita, execução, etc.)
- Confidencialidade
 - serviço que protege contra a ameaça de *intercepção de dados*, tornando os dados inteligíveis por entidades não autorizadas...

GCOM-DI-UM 77 CSI 2002, A.Costa

 **Serviços de segurança**

Universidade do Minho

- Integridade dos dados
 - serviço que oferece protecção contra a ameaça de *manipulação*, detectando quaisquer alterações...
- Não repúdio
 - prova de entrega
 - impossibilitar que o destinatário negue ter recebido
 - não repúdio da origem
 - impossibilitar que quem originou determinada informação venha a negar tal facto...

GCOM-DI-UM 78 CSI 2002, A.Costa

Universidade do Minho

Criptografia

- O exemplo de Júlio César:
 - mensagem original: "cartago esta no papo"
 - mensagem codificada: "ebtciqguxcpqrq"
 - uso de um algoritmo criptográfico baseado na substituição de letras (2 posições no alfabeto)
- Os algoritmos actuais:
 - algoritmos muito mais complexos, que dissimulam qualquer padrão existente no texto
 - número de "chaves" muito grande, que invalidem a procura por tentativas...
 - não há necessidade de manter os algoritmos secretos, mas apenas as chaves...

GCOM-DI-UM
79
CSI 2002, A.Costa

Universidade do Minho

Componentes dos Sistemas criptográficos

GCOM-DI-UM
80
CSI 2002, A.Costa

Universidade do Minho

Criptografia simétrica e assimétrica

- Sistemas simétricos:
 - chave usada na cifragem é igual á chave usada na decifragem...
 - obriga os interlocutores a negociarem uma chave antes de iniciarem uma comunicação: **Problema!**
- Sistemas assimétricos (ou de chave pública)
 - cada utilizador possui um par de chaves:
 - o texto cifrado com uma só pode ser decifrado com a outra
 - conhecendo uma delas é impossível descobrir a outra
 - utilizador mantém uma secreta (*chave privada*) e divulga a outra (*chave pública*)
 - não há necessidade de negociar nada á partida
 - Problema:** são demasiado lentos
- São muitas vezes usados em conjunto

GCOM-DI-UM
81
CSI 2002, A.Costa

Universidade do Minho

Funções de sumariação

- A partir de texto de qualquer tamanho, é produzido um sumário de tamanho fixo.
- A partir de um sumário é impossível determinar o texto que lhe deu origem
- Deve ser impossível encontrar dois textos que produzam o mesmo sumário...
- Os sumários funcionam como se fossem a impressão digital do texto sumariado

GCOM-DI-UM
82
CSI 2002, A.Costa

Universidade do Minho

Mecanismos baseados em criptografia

- Assinaturas digitais
 - tal como uma assinatura vulgar, adiciona-se ao texto para:
 - possibilitar aos destinatários a verificação da origem
 - associar o texto ao seu originador de forma a que este não possa negar esse facto
 - garantir a sua *integridade* (a mensagem recebida é realmente a mensagem que foi originada)

GCOM-DI-UM
83
CSI 2002, A.Costa

Universidade do Minho

Geração de uma assinatura digital

- produzir um sumário
- cifrar o sumário com a sua *chave privada*
- enviar a mensagem com a assinatura anexada

GCOM-DI-UM
84
CSI 2002, A.Costa

Verificação duma assinatura digital

- produzir um sumário a partir do texto original
- decifrar o sumário com a *chave pública* do originador
- comparar os dois sumários obtidos

GCOM-DI-UM 85 CSI 2002, A.Costa

Mecanismos baseados em criptografia

- Envelope digital
 - tal como os envelopes normais impede os observadores não autorizados de aceder ao conteúdo
 - realiza o serviço de *confidencialidade*
 - é usada criptografia simétrica e assimétrica em conjunto:
 - gera-se aleatoriamente uma chave secreta
 - cifra-se toda a mensagem com essa chave (algoritmo simétrico)
 - cifra-se a chave usada no ponto anterior com a chave pública do destinatário (algoritmo assimétrico)
 - envia-se todo o conjunto para o destinatário...

GCOM-DI-UM 86 CSI 2002, A.Costa

Geração de um envelope digital

GCOM-DI-UM 87 CSI 2002, A.Costa

Abrir um envelope digital

- decifrar a chave secreta usando a sua chave privada
- decifrar a mensagem usando a chave que acabou de obter

GCOM-DI-UM 88 CSI 2002, A.Costa

Mecanismos baseados em criptografia

- Certificação digital
 - Todos os mecanismos baseados em técnicas criptográficas assimétricas dependem da autenticidade das chaves públicas!
 - É preciso garantir que as chaves pertencem mesmo aos seus donos...
 - Como?
 - gerando um certificado com o nome do utilizador, a sua chave pública, a validade da chave, etc e assinando essa chave por uma *autoridade de certificação*
 - é preciso certificar as chaves dessas autoridades de certificação, formando uma cadeia de certificação

GCOM-DI-UM 89 CSI 2002, A.Costa

Exemplo de utilização no E-Mail

- Existem duas soluções para correio seguro na Internet: o PEM e o PGP
- O PGP (*Pretty Good Privacy*) é o mais popular e existe para todas as plataformas
- Passos a seguir:
 - obter o software PGP
 - criar um par de chaves para si próprio
 - divulgar a sua chave pública (de todas as formas)
 - assinar mensagens e mandar mensagens secretas!

GCOM-DI-UM 90 CSI 2002, A.Costa

Mensagem assinada

Universidade do Minho

GCOM-DI-UM 91 CSI 2002, A.Costa

Exemplo de uma chave pública

```

-----BEGIN PGP PUBLIC KEY BLOCK-----
Version: 2.6.1

mQCNAi5d64AAAEAPBjQkFCTDxvVZ6Z7q9QSWcoFbY4lyGkYaAfo0LjaBzQ
ax/BTTwabfTSDUUX8e3YalQ56wYDFomZ0Ia3a532/QPQ7Rv484eVkmhZ
+54tXxSD3obLomShaTW8axXUx6WGAsqD0vws8rHYAOwP0GkIrvsERAUR
iCZk8NIEI8bnVibCBWYWalbnNhdDsqbXAZGkadW1pbmbLaiB0PokAQIFEC59
alQbnXK777BEQEbus8EALScpOZQThwVkerCepGmMf45JfSf3f3grPEP0yQp
6qOMZa86s0L5eV1y5YerY3+EbuaVjDuarVjPHKXZZ7F9+9A3KkKCP3Bhg
mGKL5QYKFSgQkL+27ofswCqgDYP+PGW0b+W+Lsvhb8mva288XiaYZWM3
iQCVaUQLm21y7DGjgR+WpAQGUWAQAvYXkvaZ8Ylp124d20SjB1EajEzdv7
7FVtKcz6NvBdzWloBrTogp8PDLk1b+LHKBLG8F1UMFvTF0Xa23ZwIXSV
PpXldqZU+PEa3VLX84CkoCJEp+BXcLGIwAiy+N30sjYjNjmaAgBdH2hd
nB+0sVC9Ac=
-----END PGP PUBLIC KEY BLOCK-----

```

Universidade do Minho

GCOM-DI-UM 92 CSI 2002, A.Costa

World Wide Web

Universidade do Minho

GCOM-DI-UM CSI 2002, A.Costa

Hipertexto e Hipermedia

- Um conceito para:
 - a *apresentação* (forma como o documento é exibido)
 - a *acesso* (forma de aceder e obter os docs.)
 - a *estrutura* (forma com está estruturado)
 - a *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.

Universidade do Minho

GCOM-DI-UM 94 CSI 2002, A.Costa

Hipertexto e Hipermedia

- Modelo teórico: nós com ligações entre si

Universidade do Minho

■ Âncora (origem)
■ Âncora (destino)
→ Ligação

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?)

GCOM-DI-UM 95 CSI 2002, A.Costa

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*!

Os documentos hipertexto contém palavras seleccionadas também designadas por *ligações*.

Hipertexto é texto com ligações!

A forma como a expansão é provocada, depende do interface:

- clicar na âncora (interfaces gráficos!)
- digitar um número de referência (interfaces não-gráficos, só de texto)

As âncoras podem ser expandidas a qualquer altura por quem lê, para obter informações adicionais.

Universidade do Minho

GCOM-DI-UM 96 CSI 2002, A.Costa

Universidade do Minho

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.

GCOM-DI-UM CSI 2002, A.Costa

Universidade do Minho

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?

1. Documento inicial (Home)
2. Caminho percorrido é caminho de retorno

GCOM-DI-UM 98 CSI 2002, A.Costa

Universidade do Minho

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!

GCOM-DI-UM 99 CSI 2002, A.Costa

Universidade do Minho

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!

GCOM-DI-UM 100 CSI 2002, A.Costa

Universidade do Minho

World Wide Web

O que é?

- Nasce em 1989, CERN, Suíça...
- A designação é a do projecto que lhe deu origem...
- Concretiza o conceito de hipermedia na Internet

Definição → Sistema de informação *hipermedia, cooperativo, distribuído e heterogeneo*

Outros significados:

- designa o conjunto dos *protocolos* desenvolvidos pelo projecto
- designa o *hiperespaço* de informação disponível na Internet e suportado por *servidores de informação* interligados...

GCOM-DI-UM 101 CSI 2002, A.Costa

Universidade do Minho

World Wide Web

- Ideias base (revolucionárias?) do WWW:
 - Todos podem criar documentos e inseri-los na teia!**
 - qualquer um pode ser autor, incluir referências para qualquer outro documento, e ser referenciado livremente...
 - não há nenhuma autoridade administrativa, centralizante, ou restritiva de qualquer outra forma...
 - Uma forma uniforme de localizar documentos em todo o mundo!**
 - foram introduzidos métodos para definir com exactidão *onde* e *como* encontrar documentos na Internet...
 - estas *moradas*, designam-se por *Uniform Resource Locators* (URLs)
 - não são apenas *endereços*, porque incluem *métodos de acesso*

GCOM-DI-UM 102 CSI 2002, A.Costa

World Wide Web

3. Um interface com o utilizador, único e uniforme!

- similaridade de operação entre clientes distintos...
- transparência da localização (não interessa onde está!)
- esconder incompatibilidades entre sistemas na representação dos dados...

GCOM-DI-UM 103 CSI 2002, A.Costa

World Wide Web

4. Acesso a qualquer "base de dados" de informação

- é preciso garantir acesso a dados que não tenham sido produzidos especificamente para o WWW...
 - mesmo que forma mais limitada, e sem links
- são necessárias aplicações que sirvam de intermediários no acesso:
 - designam-se por *gateways de aplicação*

5. Um suporte para realizar transacções!

- o utilizador pode introduzir dados em formulários e enviá-los...
- ...o que permite a realização de vários tipos de transacções.
- Uma das utilizações, entre muitas, é o comércio electrónico!

GCOM-DI-UM 104 CSI 2002, A.Costa

WWW - Os protocolos

- *Uniform Resource Locators (URLs)*

Formato geral:

```
<protocolo-acesso>://<computador.organização.país>/pasta/subpasta/ficheiro.ext
```

- 1 - protocolo de acesso aos dados (HTTP, FTP, etc...)
- 2 - nome completo da máquina (ou endereço IP!)
- 3 - pastas dentro das quais se encontra o documento
- 4 - nome do documento e respectiva extensão!

Exemplo:

♦ <http://med-amsa.bu.edu/Gutenberg/Welcome.html>

GCOM-DI-UM 105 CSI 2002, A.Costa

WWW - Os protocolos

- *Uniform Resource Locators (URLs)*

Outros exemplos:

- <ftp://ftp.ci.uminho.pt/pub/README> (arquivo ftp)
- <http://marte.uminho.pt/MCS/index.html> (servidor www)
- <telnet://orpheu.ci.uminho.pt> (telnet)
- <gopher://gopher.uminho.pt/> (gopher)
- <news:pt.internet> (grupo de news)
- <mailto:cxxxxx@ci.uminho.pt> (envio de e-mail)

GCOM-DI-UM 106 CSI 2002, A.Costa

WWW - Os protocolos

- *HyperText Transfer Protocol (HTTP)*
- otimizado para hipertexto interativo
- Rápido: pedidos satisfeitos numa única interacção (ida e volta)

- Não orientado ao estado (*stateless*)
 - Não se estabelece nenhuma sessão entre servidor e cliente
 - O URL é suficiente para aceder ao documento
- Extensível: suporta transferência de qualquer tipo de dados
- Não é indispensável: podem ser usados outros protocolos...

GCOM-DI-UM 107 CSI 2002, A.Costa

WWW - Os protocolos

- *HyperText Transfer Protocol (HTTP)*

```
HTTP/1.0 200 OK
Server: Netscape-FastTrack/2.01
Date: Mon, 11 Feb 2002 19:49:32 GMT
Accept-ranges: bytes
Last-modified: Mon, 08 Feb 1999 15:04:21 GMT
Content-length: 969
Content-type: text/html
<-HTML>
..... Página HTML pedida ...
<-HTML>
```

GCOM-DI-UM 108 CSI 2002, A.Costa

Universidade do Minho

World Wide Web

- Problema da sobrecarga da rede
 - Assume-se uma rede com cobertura mundial e capacidades infinitas de transferência que não existem!
 - Os utilizadores nem sempre são avisados do tamanho do documento antes da transferência
 - Os utilizadores e autores não estão conscientes das limitações da rede
 - O mesmo documento pode ser transferido mais do que uma vez pelo mesmo cliente:
 - Proxy, Caching...e Mirroring
 - Livre acesso a conhecimentos *versus* negócio!

GCOM-DI-UM 115 CSI 2002, A.Costa

Universidade do Minho

World Wide Web

- Caching no Cliente WWW

1. ver cache
2. não!
3. pedido
4. resposta
5. guarda cópia em cache

- o utilizador passa várias vezes pelo mesmos documentos...
- ...se os documentos mais recentes forem guardados em cache,
- podem posteriormente (dias) ser prontamente acedidos em disco
- contribui-se para diminuir a carga na rede e o tempo de acesso

GCOM-DI-UM 116 CSI 2002, A.Costa

Universidade do Minho

World Wide Web

- Será possível ter uma cache partilhada por todos os clientes de uma organização?

Caching num servidor próprio - **Proxy** - usado por todos

GCOM-DI-UM 117 CSI 2002, A.Costa

Universidade do Minho

World Wide Web

- Para que o *caching* seja possível ao nível da organização é necessário que os pedidos de todos clientes passem por um mesmo servidor - o servidor *proxy*
- Ao receber um pedido, o *proxy* encaminha-o ao servidor original e remete a resposta obtida ao cliente, mantendo uma cópia na *cache*
- Quando é pedido um documento repetido devolve ao cliente a cópia em *cache*
- Ao configurar o seu *browser* só tem vantagens em indicar qual é o *proxy* da organização
- O *proxy* também pode permitir a *browsers* colocados em redes sem conectividade Internet aceder a documentos sem quaisquer restrições

GCOM-DI-UM 118 CSI 2002, A.Costa

Universidade do Minho

World Wide Web

- Proxy e Caching

GCOM-DI-UM 119 CSI 2002, A.Costa

Universidade do Minho

World Wide Web

- Mirroring

- Objectivos:
 - Disponibilizar *réplicas* mais próximo dos utilizadores, optimizando a utilização dos recursos da rede
 - Distribuição de "carga" (nº pedidos) entre servidores...
 - Tolerância a faltas: se um não estiver disponível, há substitutos

GCOM-DI-UM 120 CSI 2002, A.Costa

World Wide Web

• *Robots* ou *Spiders*

Estes clientes especiais, surgem para tentar realizar tarefas que com o crescimento do WWW deixaram de poder ser feitas manualmente...

São **Agentes de Software** que exploram autonomamente alguma porção da Web, seguindo todas as ligações (*links*) existentes!

→

- Os percursos cíclicos são evitados...
- A porção a ser explorada pode ser demarcada por:
 - *domínios* ou *sites*
 - nº de *níveis*
 - combinações dos dois anteriores

GCOM-DI-UM 121 CSI 2002, A.Costa

World Wide Web

• *Robots* - Utilização

Todos os *Robots* realizam as seguintes tarefas, continuamente:

- Obter página...
- Procurar e recolher todos os URLs contidos nessa página
- Fazer algum tipo de processamento
- Próxima página ainda não visitada...

É no processamento que fazem que os *Robots* se distinguem entre si

- Ex: **Robot de replicação** (*mirror*) cuja tarefa é gravar todas as páginas de um *site* no disco local, corrigindo todas as referências para continuarem válidas localmente!...

GCOM-DI-UM 122 CSI 2002, A.Costa

World Wide Web

• Além da **replicação**, as tarefas mais comuns são:

Estatísticas WWW

- Recolher dados estatísticos: nº de *sites*, documentos por *site*, tamanho médio dos documentos, nº de imagens por pág., etc.

Manutenção

- Testar a integridade das ligações internas e externas (apenas os primeiros links para fora) e alertar para grandes alterações realizadas em documentos. Auxiliam a gestão de um *site*...

Construção de Índices

- Criar enormes bases de dados que indexam os documentos e ficheiros existentes no WWW. Alguns catalogam todo o texto, outros apenas títulos ou resumos... São os mais populares!

GCOM-DI-UM 123 CSI 2002, A.Costa

World Wide Web

Os chamados **Agentes Inteligentes**, são os novos robots pessoais que podem ser "treinados" pelo seu dono (assuntos de interesse) e depois "buscam" informações semelhantes na rede. Os mais sofisticados usam mesmo técnicas de Inteligência Artificial...

Problemas com os Robots:

- "inundam" os servidores com pedidos
 - disparam pedidos muito rapidamente e por vezes em paralelo!
- "entopem" a rede
 - consomem largura de banda, prejudicando todos (sites visitados e também a rede local!)

• **Soluções:**

- ficheiro *robots.txt* contendo o nome dos indesejados!...
- detectar e impedir acessos em massa da mesma origem!...

GCOM-DI-UM 124 CSI 2002, A.Costa

WWW - Pesquisa avançada

"É muito fácil pesquisar... o difícil é pesquisar bem!..."

• A importância crescente do "**Information Retrieval**":

- Grande número de documentos multimedia acessíveis na Internet através do WWW, FTP, etc...
- Bibliotecas digitais
- Necessidade de mecanismos eficazes e eficientes para descobrir informação relevante
- Devido a limitações na tecnologia actual:
 - texto apenas, em vez de multimedia

GCOM-DI-UM 125 CSI 2002, A.Costa

WWW - Pesquisa avançada

Ferramentas de pesquisa são programas que fornecem informação...

• **Classificação das ferramentas de pesquisa:**

Catálogos	Motores de Busca	Mistos	Multi-Engenho
Pesquisa hierárquica "por assunto"	Pesquisa "por palavras chave"...	Combinação dos anteriores	Ferramenta que interroga várias outras em paralelo

• **Ou, do ponto de vista da indexação:**

Indexação manual	Indexação total (<i>full</i>)	Indexação parcial
Feita por utilizadores, que classificam em categorias!	Indexação automática de todas as palavras!	Indexação de apenas algumas palavras chave!

GCOM-DI-UM 126 CSI 2002, A.Costa

WWW - Pesquisa avançada

Universidade do Minho

- **Modelo Booleano**
 - Base de dados consiste em listas de ocorrências de palavras:
 - lista (K1) = {D1, D2, D3, D4};
 - lista (K2) = {D1, D2};
 - lista (K3) = {D1, D2, D3};
 - lista (K4) = {D1}
 - A interrogação do utilizador é um conjunto de palavras chave ligadas por operadores booleanos:
 - Interrogação: (K1 AND K2) OR (K3 AND (NOT K4))
 - O sistema devolve todos os documentos (Di) para os quais a interrogação é verdadeira:
 - Resposta: D1, D2, D3
- Utilizado por um grande número de sistemas operacionais, mas muito difícil de usar por utilizadores ocasionais...

GCOM-DI-UM 127 CSI 2002, A.Costa

WWW - Pesquisa avançada

Universidade do Minho

- **Modelo Vectorial**
 - cada documento é representado por um vector de termos cujos pesos são o número de ocorrências no documento...
 - num espaço de dimensão T (número de termos distintos na colecção)
 - também a interrogação é vista como um vector...
 - Os vectores que representam o documento e a interrogação são confrontados e é calculado um valor *Retrieval Status Value* (RSV)
 - Para cada interrogação, os documentos são mostrados ao utilizador por ordem decrescente desse valor, podendo ser inspeccionados...
 - a relevância dos documentos ou partes inspeccionadas pelo utilizador pode servir para realimentar o sistema com novas interrogações (*realimentação de relevância*)

GCOM-DI-UM 128 CSI 2002, A.Costa

WWW - Pesquisa avançada

Universidade do Minho

- **Dicas e truques para uma boa pesquisa**
 - Adopte uma estratégia baseada naquilo que sabe:
 - Use os catálogos, se sabe em que tópicos gerais se inclui aquilo que procura...
 - Use os motores de busca (*c/* indexação total) se conhece coisas específicas como nomes ou palavras chave...
 - Características adicionais, como autor, localização, língua, etc, podem ser relevantes para auxiliar a pesquisa
 - Utilizar palavras pouco comuns e se possível diferenciadas...
 - Utilize vários motores de busca e procure sempre entender os resultados, distinguindo os “falsos acertos”

GCOM-DI-UM 129 CSI 2002, A.Costa

WWW - Pesquisa avançada

Universidade do Minho

- **Utilização de alguns operadores para construir perguntas mais sofisticadas:**
 - Operadores lógicos (AND, OR, NOT) e de proximidade (NEAR)
 - Sinais mais (+) e menos (-) para incluir ou excluir termos
 - Entre aspas (“...”) para definir uma frase exacta ou palavras muito relacionadas que devem ser procuradas em conjunto...
 - Parentesis rectos [..., ...] para agrupar palavras a serem pesquisadas quer isoladamente quer em conjunto
 - Operadores “mais como este” ou “more like this” (pesquisas por exemplos)
 - Letras maiúsculas, usadas para definir nomes próprios ou acrónimos...
- Nem todos os operadores são suportados por todos os serviços de pesquisa!

GCOM-DI-UM 130 CSI 2002, A.Costa

WWW - Pesquisa avançada

Universidade do Minho

- **Exemplo 1:** financiamento de investigações sobre o cancro
 - utilizando operadores lógicos para definir sinónimos:
 - (*cancer OR oncol**) AND *research* AND (*grant* OR fund**)
 - utilizando parentesis, ou aspas:
 - [*cancer, oncol, research, grant, fund*]
 - “*cancer research grant*”
 - operadores de proximidade
 - cancer NEAR research NEAR grant*
- **Exemplo 2:** informação sobre Warner-Lambert ou sobre o seu grupo de investigação farmacêutica Parke-Davis
 - Letras maiúsculas para nomes próprios e aspas para agrupar:
 - “*Parke Davis*” OR “*Warner Lambert*”
 - “*Warner-Lambert*”, “*Parker-Davis*”

GCOM-DI-UM 131 CSI 2002, A.Costa