

# World Wide Web

---

Sérgio Nunes

Comunicações Digitais e Internet  
Ciências da Comunicação, U.Porto 2012/13

# Objetivos

---

- Definir e descrever o que é a World Wide Web.
- Conhecer as origens e as tecnologias base que suportam a web.
- Compreender como se identificam os recursos publicados na web.
- Compreender como são representados e interligados os documentos web.
- Adquirir uma visão global dos principais fluxos de informação na web.

# Origens e Tecnologias da Web

# World Wide Web

---

- **O que é a World Wide Web?**
- É um serviço de informação que funciona sobre a Internet e assenta numa estrutura baseada em documentos interligados.
- Aspectos importantes:
  - Serviço de informação.
  - Funciona sobre a Internet.
  - Documentos hipertexto interligados.

# Internet

---

- **A Internet é uma rede de dimensão mundial que resulta da interligação de diferentes redes de computadores — é uma rede de redes.**
- A Internet permite que qualquer computador ligado à rede comunique diretamente com qualquer outro computador também ligado à rede.
- A **interoperabilidade** é uma das principais características da Internet. A comunicação é possível entre computadores ligados à Internet independentemente de marcas, modelos ou formas de acesso à rede.
- A interoperabilidade é possível devido ao equipamento que permite ligar redes diferentes (os routers) e os programas que definem como se processa a comunicação (os protocolos).

# Ligação à Internet

---

- Os utilizadores acedem à Internet através de um fornecedor de acesso à Internet (do inglês Internet Service Provider — ISP), público ou privado.
- Os meios de ligação à Internet mais comuns são a rede telefónica fixa (analógica ou digital), a rede telefónica móvel, o cabo (televisão), o satélite ou o cabo de fibra ótica.
- A escolha de um ISP depende de vários fatores, em particular depende da disponibilidade geográfica, dos serviços oferecidos, dos limites fixados, da qualidade do serviço e do custo final.

# Cientes e Servidores

---

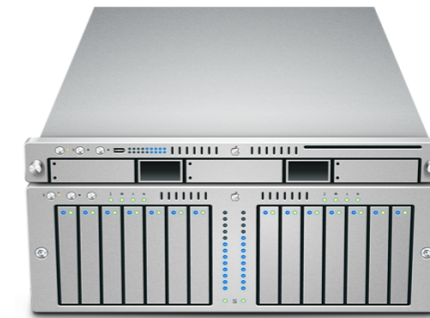
- Uma rede de comunicação não tem capacidade de processamento de informação, apenas transfere os dados de um ponto para outro ponto.
- Qualquer troca de informação através da rede é feita em pares e funciona segundo uma organização do tipo **cliente-servidor**. Os clientes pedem informação aos servidores, os servidores enviam a resposta para o cliente.
- Os clientes são os programas que iniciam os pedidos de informação.
- Os servidores são os programas que respondem aos pedidos de informação. Tipicamente, um servidor apenas consegue responder a milhares de pedidos em cada segundo.

# Interação Cliente-Servidor

---



Cliente



Servidor

**1.** O servidor aguarda o contacto de um cliente.

**2.** O cliente solicita uma ligação ao servidor.



**3.** O servidor recebe o contacto e a ligação é estabelecida.



**4.** As mensagens são trocadas entre o cliente e o servidor.



**5.** Concluída a troca de dados, as ligações são encerradas.





# Servidores de Correio Eletrónico

---



Cliente (p.e. Outlook)

**1.** O servidor aguarda o contacto de um cliente.

**2.** O cliente pede as mensagens da caixa abc@servidor



**3.** O servidor envia as novas mensagens para o cliente.



**4.** A ligação entre o cliente e o servidor é encerrada.



Servidor de Correio Eletrónico

# Internet e Web

---

- **Qual a diferença entre Internet e a World Wide Web?**
- A Internet fornece a interligação dos computadores. Transporta informação.
- A World Wide Web é um sistema de documentos hipertexto interligados.
- A Internet transporta os documentos que existem na World Wide Web.
- A World Wide Web é um dos vários serviços que funciona sobre a Internet. Outros serviços são: correio eletrónico, mensagens instantâneas, partilha de ficheiros em P2P, vídeo conferência, etc.

# Origens

---

- No final dos anos 80, um grupo de investigadores do Centro Europeu para a Investigação Nuclear (CERN) iniciou o desenvolvimento de um sistema para permitir partilha de documentos científicos.
- A troca de informação através da Internet (que já existia desde os anos 70) apresentava várias dificuldades: diferentes formatos de ficheiros, dificuldades na divulgação do trabalho — envio caso a caso.
- Principais necessidades dos investigadores:
  - Aceder remotamente aos trabalhos.
  - Aceder de forma independente do sistema operativo e dos programas.
  - Divulgar e partilhar documentos pela comunidade.

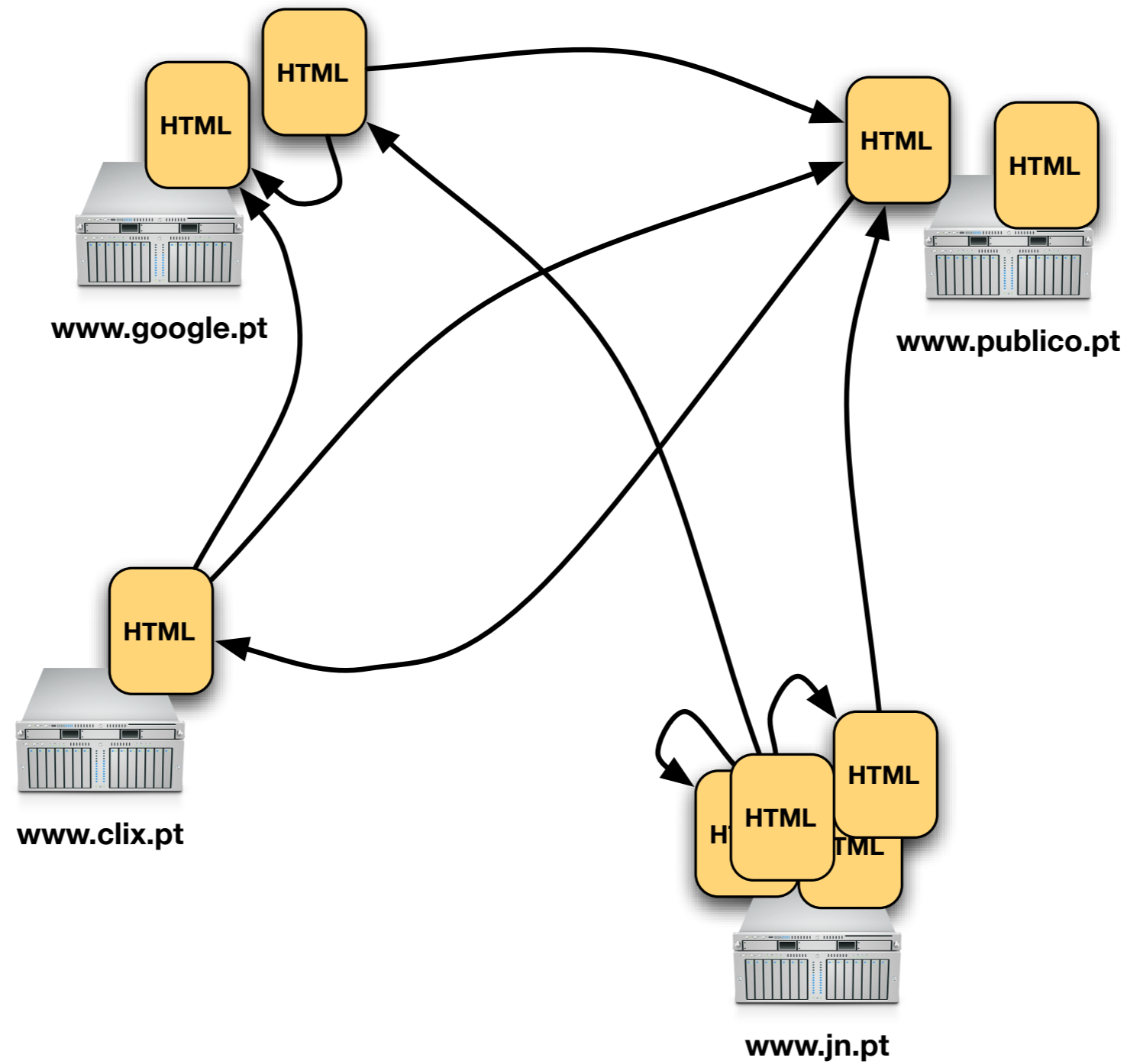
# World Wide Web

---

- A solução proposta em 1989 por Tim Berners-Lee e Robert Cailliau foi a construção de um sistema de hipertexto distribuído em rede.
- Aspectos chave:
  - Documentos distribuídos pela rede.
  - Interligação de documentos e recursos (p.e. imagens).
  - Formato universal para os documentos.
- A WWW ultrapassou largamente o âmbito para o qual foi originalmente desenhada — partilha de documentos científicos. É hoje um dos principais meios de comunicação com mais de 2 mil milhões de utilizadores e mais de 250 milhões de sítios web em funcionamento.

# World Wide Web

---



# Hipertexto

---

- "Hipertexto é uma técnica de armazenamento e apresentação de informação baseada num sistema de referências cruzadas que formam uma rede de associações." (Texto Editora)
- Com hipertexto é possível definir estruturas não lineares construídas com base na utilização de referências cruzadas entre textos.
- Este conceito foi descrito pela primeira vez em 1945 por Vannevar Bush no ensaio "As We May Think".
- O termo **hypertext** foi proposto por Ted Nelson em 1965.

# Hipertexto

As armas e os barões assinalados,  
Que da ocidental praia Lusitana,  
Por mares nunca de antes navegados,  
Passaram ainda além da Taprobana,  
Em perigos e guerras esforçados,  
Mais do que prometia a força humana,  
E entre gente remota edificaram  
Novo Reino, que tanto sublimaram;

E também as memórias gloriosas  
Daqueles Reis, que foram dilatando  
A Fé, o Império, e as terras viciosas  
De África e de Ásia andaram devastando;  
E aqueles, que por obras valerosas  
Se vão da lei da morte libertando;  
Cantando espalharei por toda parte,  
Se a tanto me ajudar o engenho e arte.

Cessem do sábio Grego e do Troiano  
As navegações grandes que fizeram;  
Cale-se de Alexandro e de Trajano  
A fama das vitórias que tiveram;  
Que eu canto o peito ilustre Lusitano,  
A quem Neptuno e Marte obedeceram:  
Cesse tudo o que a Musa antiga canta,  
Que outro valor mais alto se alevanta.

E vós, Tágides minhas, pois criado  
Tendes em mim um novo engenho ardente,  
Se sempre em verso humilde celebrado  
Foi de mim vosso rio alegremente,  
Dai-me agora um som alto e sublimado,  
Um estilo grandiloquo e corrente,  
Porque de vossas águas, Febo ordene  
Que não tenham inveja às de Hipoerene.

Dai-me uma fúria grande e sonora,  
E não de agreste avena ou frauta ruda,  
Mas de tuba canora e belicosa,  
Que o peito acende e a cor ao gesto muda;  
Dai-me igual canto aos feitos da famosa  
Gente vossa, que a Marte tanto ajuda;  
Que se espalhe e se cante no universo,  
Se tão sublime preço cabe em verso.

As armas e os barões assinalados,  
Que da ocidental praia Lusitana,  
Por mares nunca de antes navegados,  
Passaram ainda além da Taprobana,  
Em perigos e guerras esforçados,  
Mais do que prometia a força humana,  
E entre gente remota edificaram  
Novo Reino, que tanto sublimaram;

E vós, Tágides minhas, pois criado  
Tendes em mim um novo engenho ardente,  
Se sempre em verso humilde celebrado  
Foi de mim vosso rio alegremente,  
Dai-me agora um som alto e sublimado,  
Um estilo grandiloquo e corrente,  
Porque de vossas águas, Febo ordene  
Que não tenham inveja às de Hipoerene.

Dai-me uma fúria grande e sonora,  
E não de agreste avena ou frauta ruda,  
Mas de tuba canora e belicosa,  
Que o peito acende e a cor ao gesto muda;  
Dai-me igual canto aos feitos da famosa  
Gente vossa, que a Marte tanto ajuda;  
Que se espalhe e se cante no universo,  
Se tão sublime preço cabe em verso.

E também as memórias gloriosas  
Daqueles Reis, que foram dilatando  
A Fé, o Império, e as terras viciosas  
De África e de Ásia andaram devastando;  
E aqueles, que por obras valerosas  
Se vão da lei da morte libertando;  
Cantando espalharei por toda parte,  
Se a tanto me ajudar o engenho e arte.

Cessem do sábio Grego e do Troiano  
As navegações grandes que fizeram;  
Cale-se de Alexandro e de Trajano  
A fama das vitórias que tiveram;  
Que eu canto o peito ilustre Lusitano,  
A quem Neptuno e Marte obedeceram:  
Cesse tudo o que a Musa antiga canta,  
Que outro valor mais alto se alevanta.

- Estrutura fragmentada e não linear.
- A sequência da leitura depende do leitor, não do produtor do conteúdo.
- Acesso direto às partes.

# Servidor Web

---

- Um servidor web é um computador ligado à Internet onde documentos e outros recursos web estão armazenados.
- Os servidores web estão em funcionamento permanente a aguardar pedidos por parte dos clientes. Quando um pedido é recebido, o servidor prepara o documento e envia-o para o cliente.
- Exemplos de pedidos: "obter o documento x", "obter a imagem y".
- Existem milhões de servidores web em funcionamento. Cada servidor web armazena alguns documentos e recursos web. Um único sítio web pode estar disperso por vários servidores.



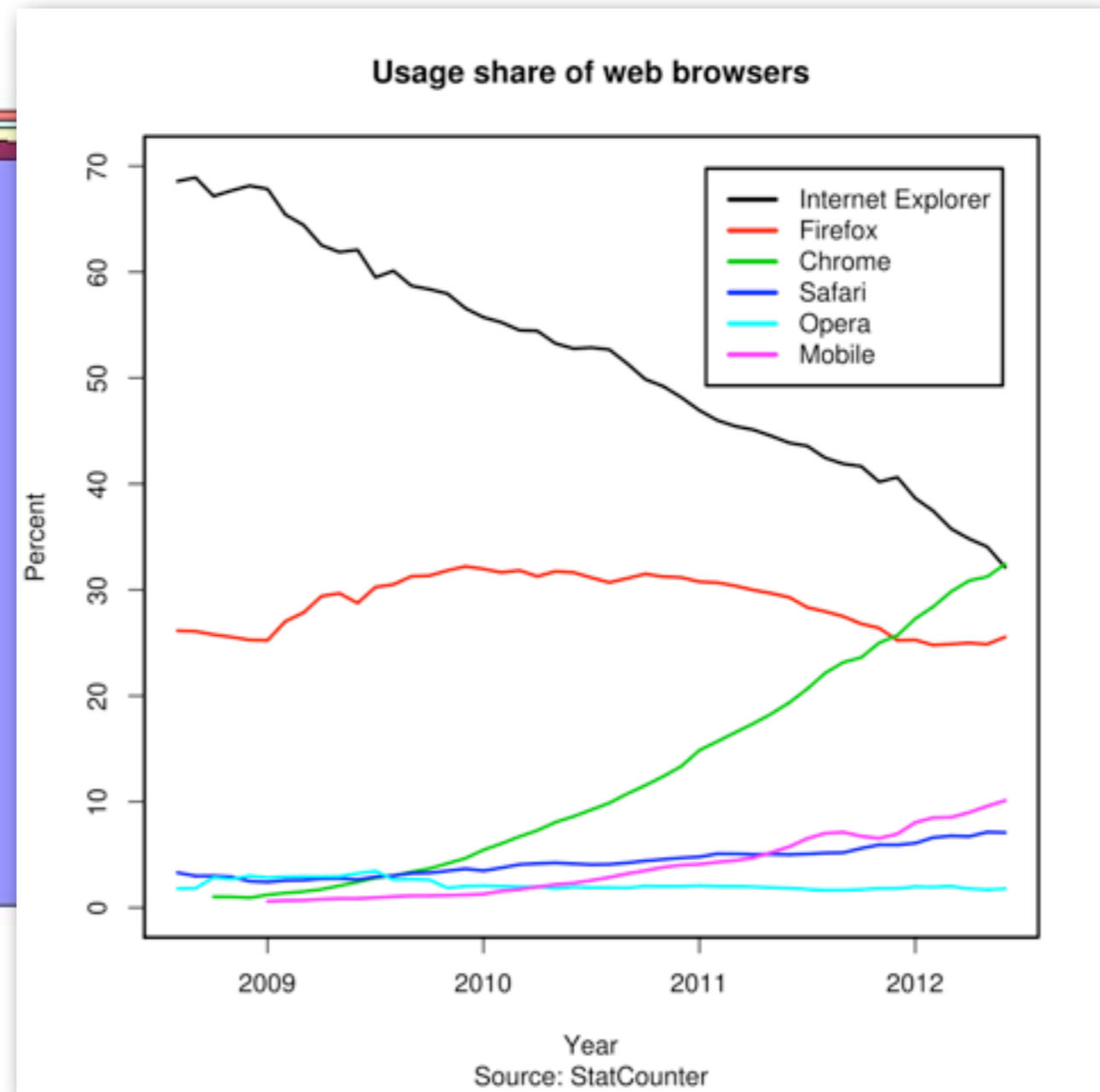
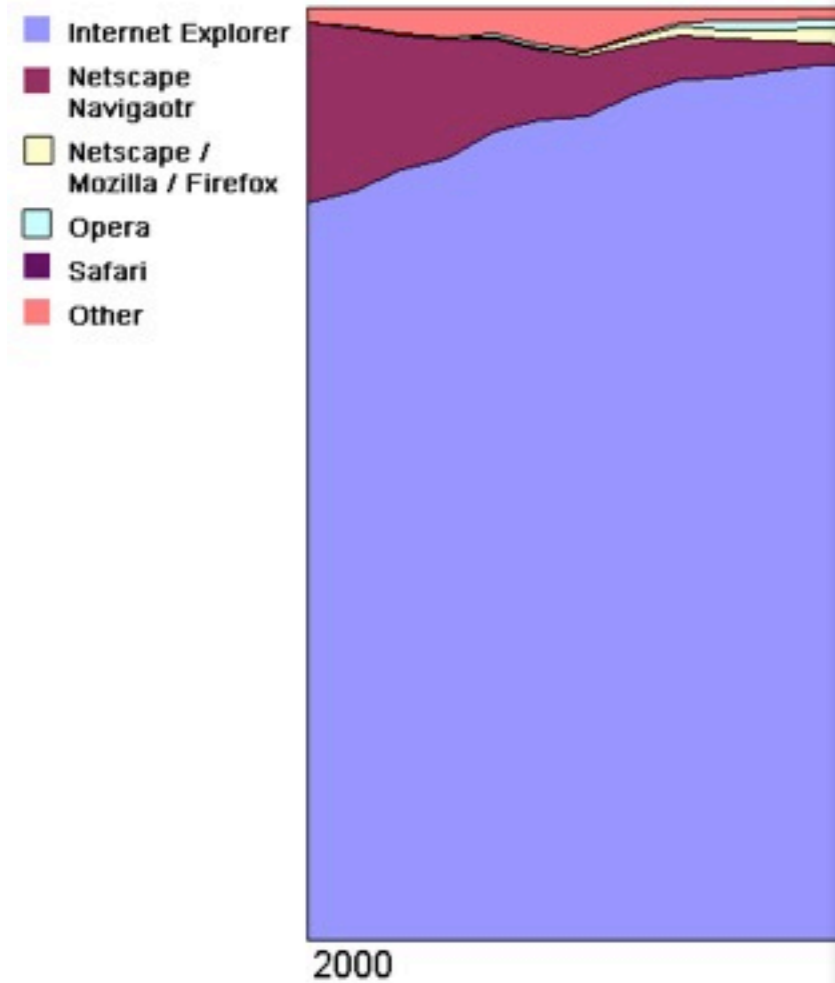
# Navegador Web

---

- **O que é um navegador web ou browser?**
- Um inquérito de rua: <http://youtu.be/o4MwTvtYrUQ>.
- Um navegador web é um programa de computador, usado para aceder e navegar em recursos disponíveis na World Wide Web.
- Sinónimos: Navegador Web, Web Browser, Browser.
- Principais opções: Internet Explorer, Firefox, Google Chrome, Apple Safari.



# Popularidade dos Navegadores



Fonte: [http://en.wikipedia.org/wiki/Usage\\_share\\_of\\_web\\_browsers](http://en.wikipedia.org/wiki/Usage_share_of_web_browsers)

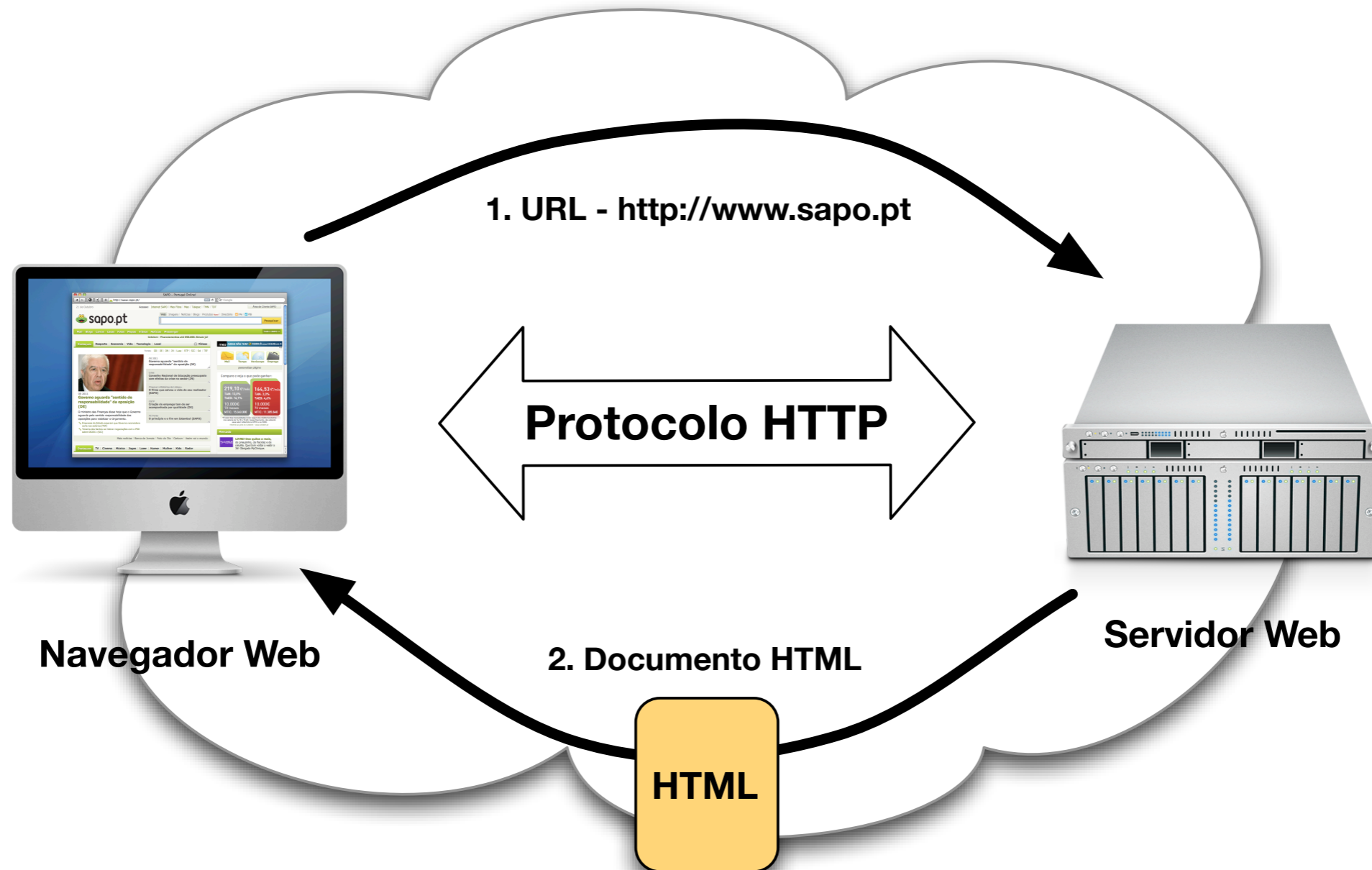
# Arquitetura da Web

---

- A World Wide Web está assente em três funcionalidades centrais:
  - **Endereçamento – URL**  
Cada documento ou recurso existente na web tem um endereço único. Os URL são usados para identificar de forma única cada documento ou recurso (imagem, vídeo, etc).
  - **Transferência – HTTP**  
Os documentos são transferidos entre o servidor web e o navegador web. A forma como se processa esta transferência é definida pelo protocolo HTTP, pertencente à família TCP/IP.
  - **Representação – HTML**  
Os documentos seguem uma estrutura padrão definida segundo a linguagem HTML.

# Arquitetura da Web

---



# Arquitetura da Web

---

- Exemplo da sequência de pedidos entre um navegador e um servidor web:
  1. O utilizador escreve um URL no navegador.
  2. O navegador envia um pedido, usando o protocolo HTTP, para o servidor indicado no endereço (URL).
  3. O servidor prepara o documento HTML indicado no pedido e, usando o protocolo HTTP, inicia o envio para o navegador.
  4. O navegador processa e apresenta o HTML que recebe. Quando o documento contém referências para outros elementos, o navegador faz automaticamente novos pedidos ao servidor.
  5. Este processo só termina quando a página está totalmente carregada.

# Identificação dos Recursos

---

- **Qualquer recurso existente na web tem um endereço único.**
- Os endereços dos recursos web são definidos com Uniform Resource Locators (URL). Os URL têm a seguinte estrutura:

**serviço://domínio/caminho/ficheiro**

- **serviço:** protocolo de comunicação a usar (http, ftp, etc).
- **domínio:** domínio a contactar.
- **caminho:** estrutura de pastas que indica o local do ficheiro.
- **ficheiros:** recurso específico pretendido.

# Identificação dos Recursos

---

- Alguns exemplos de endereços de recursos:
  - <http://www.up.pt>
  - <http://www.google.com/images/logo.gif>
  - <http://jpn.c2com.up.pt>
  - <http://jpn.c2com.up.pt/mundo/>
  - <http://jpn.c2com.up.pt/equipa.html>

# Transferência de Recursos

---

- A transferência de recursos entre os navegadores web e os servidores web é feita de acordo com o protocolo HyperText Transfer Protocol (HTTP).
- O protocolo HTTP é do tipo pedido-resposta. Cada interação entre os navegadores e os servidores web tem apenas um pedido e uma resposta. A resposta inclui o recurso pedido em caso de sucesso ou uma mensagem de erro.
- Alguns dos erros mais comuns são:
  - 403 Forbidden — Acesso ao recurso é proibido.
  - 404 Not Found — Recurso não encontrado.
  - 500 Internet Server Error — Erro interno do servidor.



# Transferência de Recursos

- O acesso a uma página web resulta em várias transferências entre o navegador web e o servidor web. Existe um pedido e uma resposta para cada recurso usado pela página.
- Uma página web inclui habitualmente dezenas de recursos.



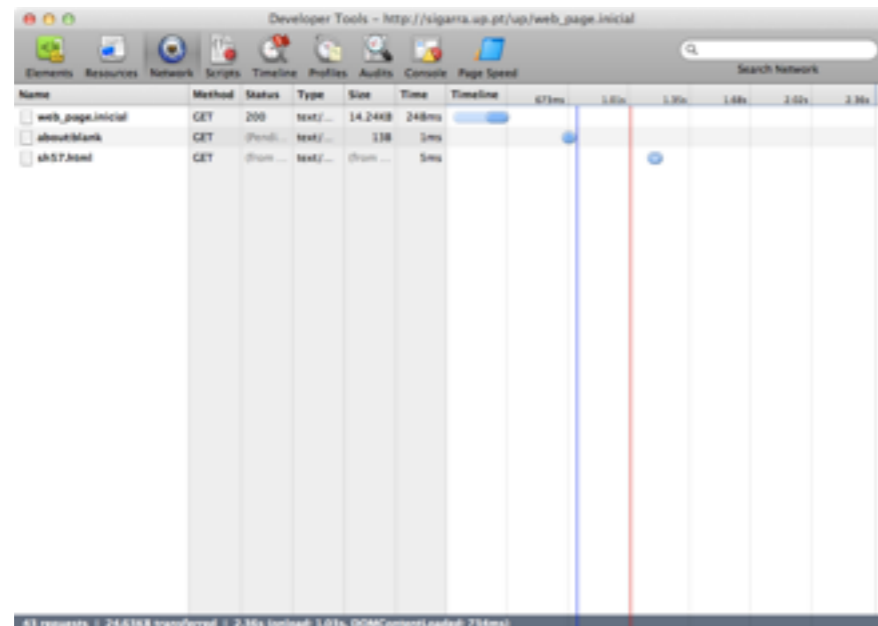
# Transferência de Recursos



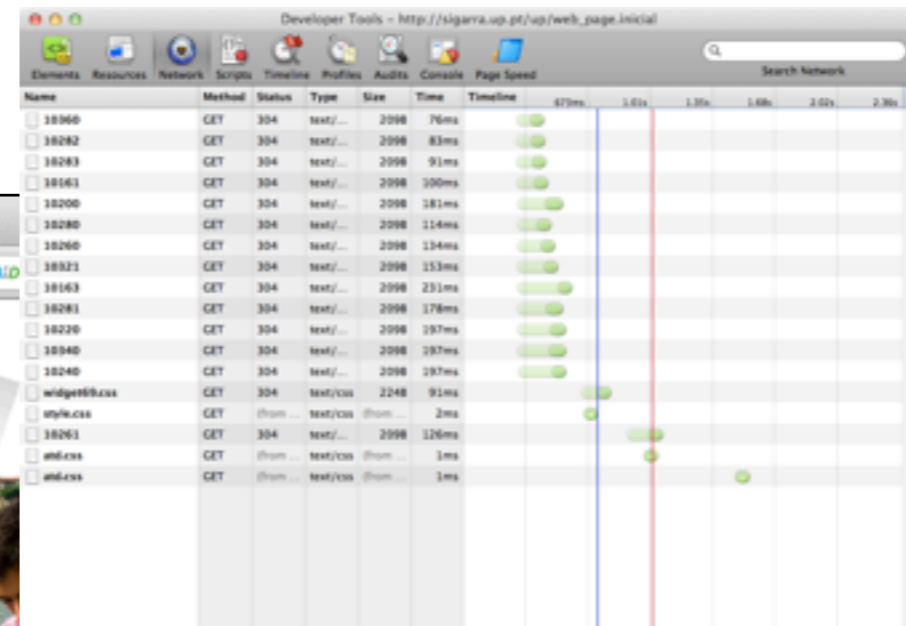
1. Documento HTML.
2. Estilos CSS.
3. Imagens.

[www.up.pt](http://www.up.pt)

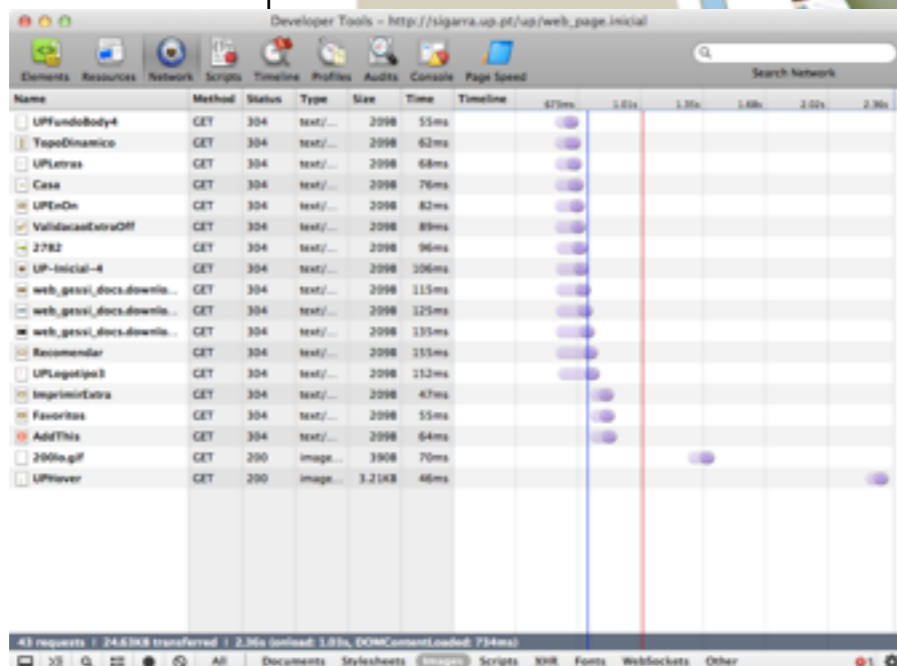
# Transferência de Recursos



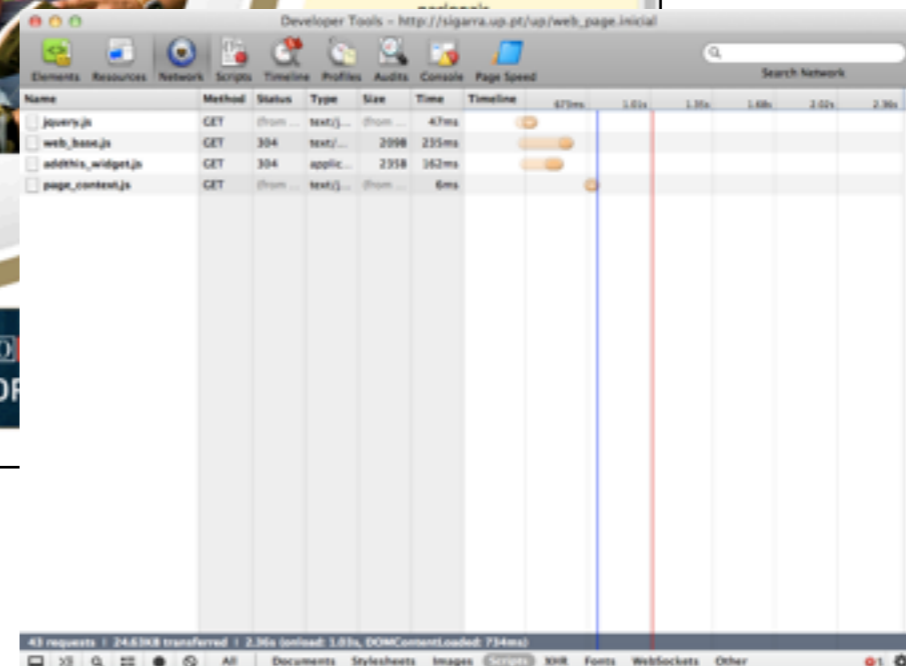
Documentos (3)



Estilos (18)



Imagens (18)



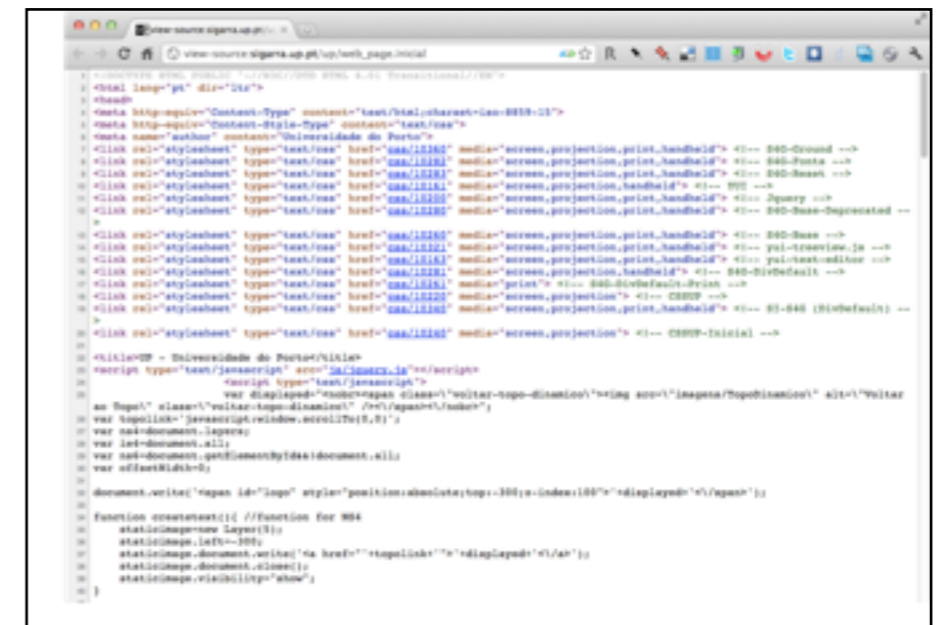
Scripts (4)

# Estrutura dos Documentos

- Os documentos web são escritos em linguagem HTML.
- A linguagem HTML permite especificar a estrutura dos documentos web e definir como se estabelecem as ligações entre os diversos documentos.
- O código HTML usado na construção de qualquer página é público e pode ser consultado usando a função de "View Source" tipicamente disponível nos navegadores web.



view source





# Estatísticas sobre a World Wide Web

---

- A web tem atualmente mais de 2 mil milhões de utilizadores.
- Existem mais de 250 milhões de sítios web em funcionamento.
- Estima-se que o total de páginas ultrapasse os 25 mil milhões.
- O número de URLs ultrapassa o bilião (um milhão de milhões =  $10^{12}$ ).
- Mais de 50% do conteúdo da web é em inglês.  
Outros idiomas populares: alemão (7,7%), francês (5,6%) e japonês (4,9%).

Fonte: Wikipédia — [http://en.wikipedia.org/wiki/World\\_Wide\\_Web#Statistics](http://en.wikipedia.org/wiki/World_Wide_Web#Statistics)

# Impacto da Web

---

- O impacto da World Wide Web ultrapassou largamente os objetivos iniciais focados na partilha de informação entre cientistas. É difícil encontrar uma área da sociedade que tenha ficado inalterada após o surgimento da web.
- Qualquer área associada à produção ou distribuição de informação sofreu alterações importantes com a adoção generalizada da web.
- Novos modelos de negócio foram criados (e destruídos). Exemplos?
- Serviços ou produtos considerados insustentáveis surgiram na web: imagens de satélite de toda a Terra, mapas das ruas, Wikipédia, informações sobre *qualquer assunto* à distância de um clique, etc.

# Sumário

---

- A World Wide Web é um serviço que funciona sobre a Internet.
- A World Wide Web é um sistema de informação baseado na interligação de documentos hipertexto dispersos por servidores web.
- A World Wide Web foi desenvolvida no CERN, na Suíça, no final dos anos 80 do século passado, por Tim Berners-Lee e Robert Cailliau para agilizar o processo de partilha de informação em ambiente científico.
- A World Wide Web funciona segundo um paradigma cliente-servidor.
- Existem três aspetos nucleares que caracterizam a arquitetura da web: a identificação dos recursos com base nos URL, a transferência dos recursos entre os servidores e os navegadores com o protocolo HTTP, e uma representação uniforme dos documentos com base na linguagem HTML.

# Impacto da Web na Difusão e Consumo de Informação



# Ideias Chave

---

- A web é um sistema de informação distribuído de alcance mundial. O impacto ao nível da difusão e consumo de informação é transversal e muito significativo.
- Ideias chave:
  - Acesso do indivíduo ao espaço público.
  - *Desintermediação* da comunicação.
  - Convergência dos meios.
  - Consumo de informação assíncrono.

# Acesso ao Espaço Público

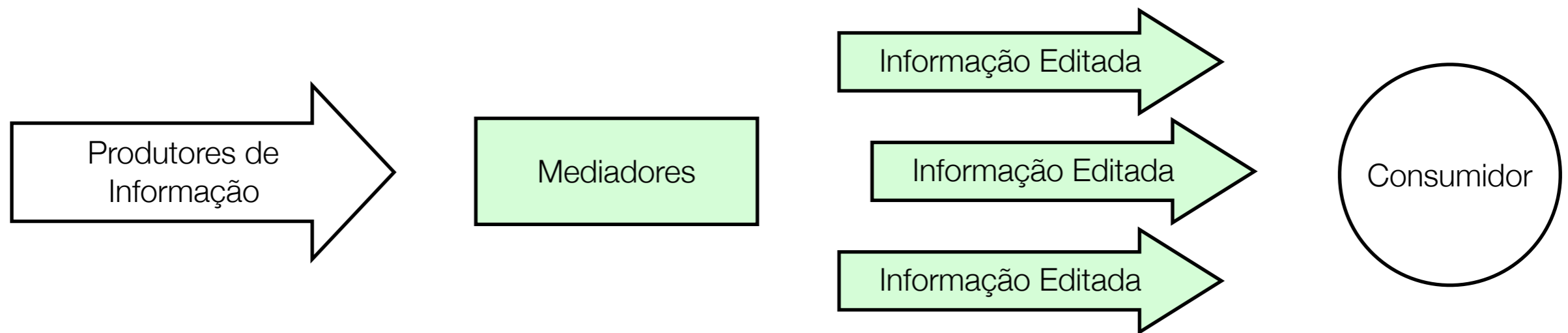
---

- **Qualquer indivíduo pode ter uma voz no espaço público.**
- A web eliminou os custos tradicionalmente associados à cópia e transporte de informação. Impacto significativo em alguns modelos de negócio.
- A disseminação de ferramentas gratuitas (p.e. blogues) eliminaram os custos de alojamento e publicação.
- Sugerem novas configurações para a publicação de informação: páginas pessoais, blogues, microblogues, redes sociais, wikis, etc.

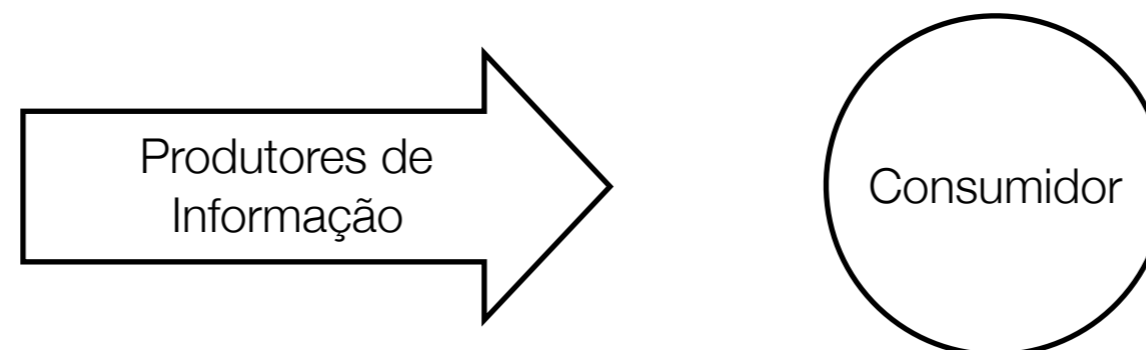
# Desintermediação da Comunicação

---

- No paradigma tradicional, o espaço público é mediado — exemplos da televisão, rádio ou imprensa.



- A web modificou o espaço público. As fontes tornaram-se diretamente acessíveis e os mediadores tradicionais perderam relevância.



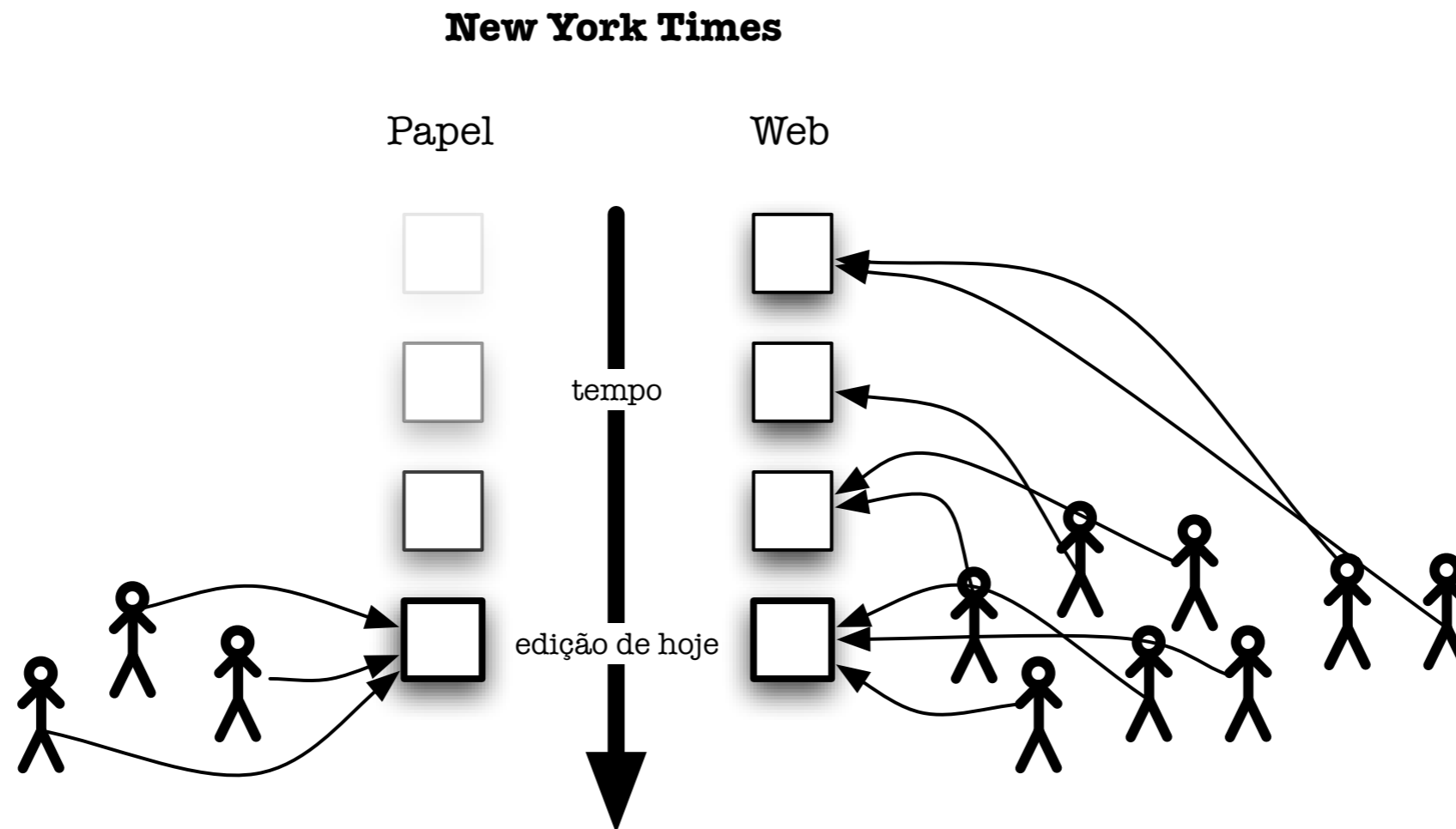
# Convergência dos Meios

- A web suporta diferentes formatos (áudio, vídeo, texto), o que tem resultado na convergência dos diferentes meios para a web — rádio, televisão, etc.



# Consumo Assíncrono

- Ao contrário dos meios tradicionais, a web permite a acumulação do conteúdo de forma permanente. Os arquivos e a possibilidade de acesso individual aos fragmentos de informação são aspectos fundamentais.



# TIME

PERSON OF THE YEAR



Yes, you.  
You control the Information Age.  
Welcome to your world.

2006

# Novo Paradigma

---

- O espaço comunicacional, anteriormente organizado e estruturado, torna-se fragmentado, disperso e dinâmico.
- As anteriores ferramentas (p.e. jornais) deixam de fazer sentido neste contexto — *porquê pagar por um conjunto limitado de notícias, impressas em papel, e que foram escritas ontem?*
- **Uma nova realidade exige novas ferramentas.**
  - A explosão no número de fontes → como gerir a quantidade?
  - O surgimento de novos produtores → como avaliar a qualidade?
  - A sobrecarga de informação → como identificar o importante?

# Novas Ferramentas | RSS

---

- **Sem intermediários, como organizar a consulta das fontes?**
- Novas tecnologias permitem a subscrição de fontes e o acompanhamento automático de alterações. A informação é entregue ao utilizador quando surge.
- Tecnologia chave: **RSS** — linguagem que permite estruturar e sistematizar alterações em conteúdos. Os conteúdos tornam-se 'processáveis' por máquinas.



# RSS | Aplicações

---

- Há um número crescente de serviços a publicar os conteúdos em formato RSS: jornais, blogues, organizações, wikipédia, motores de pesquisa, etc.
- O número de cenários possíveis para a subscrição e filtragem de conteúdos é cada vez maior.
  - Receber todos os textos de um blogue.
  - Todos os texto de uma categoria de um blogue.
  - Resultados novos da pesquisa por uma expressão.
  - Alertas de alterações a uma página.
  - Notícias publicadas por um jornal online.
  - Citações a um nome ou entidade.

# Novas Ferramentas | Editores Automáticos

- Sistemas informáticos são capazes de extrair informação de milhares de fontes e hierarquizar essa informação com base em diversos sinais.



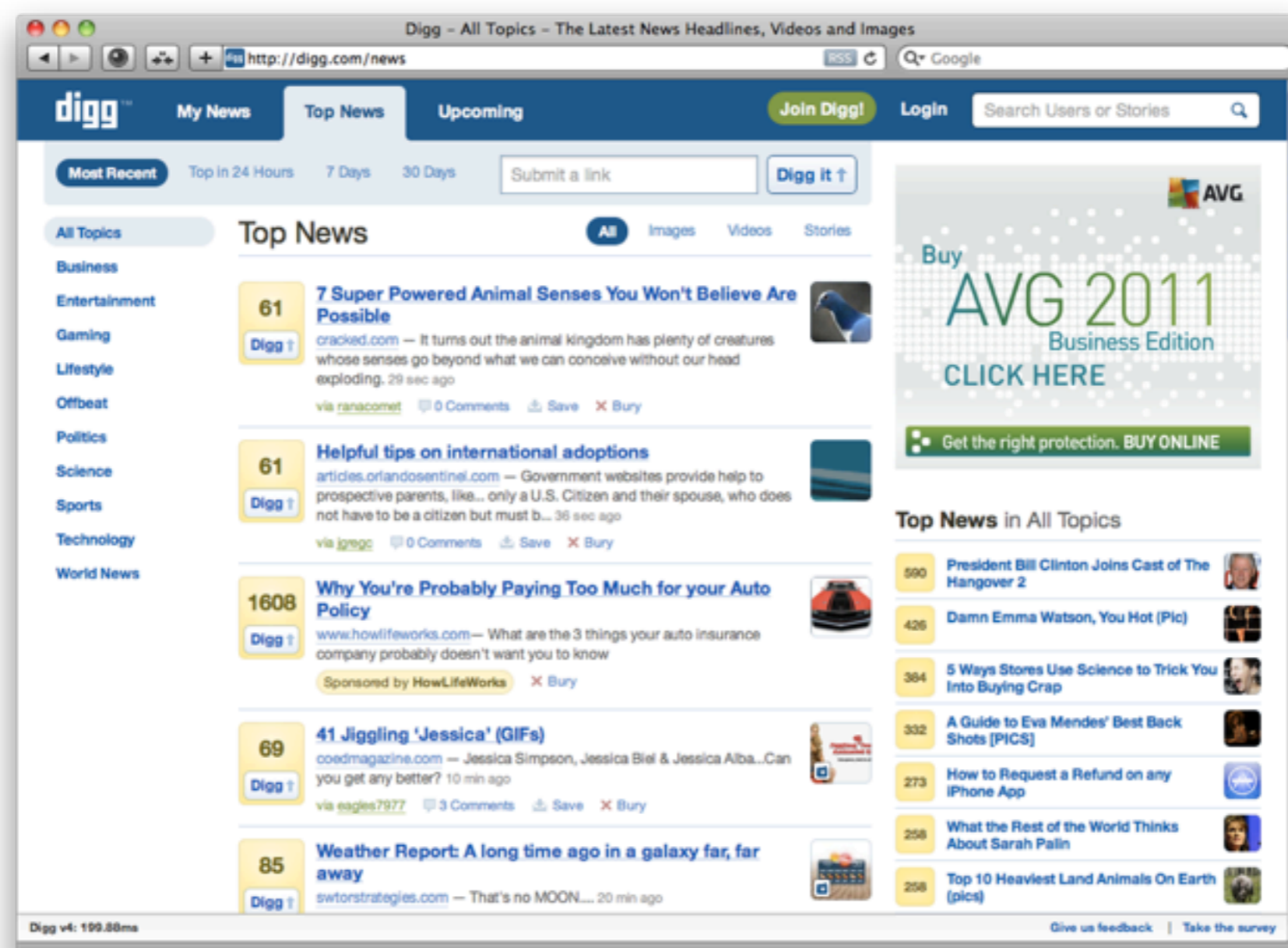
# Novas Ferramentas | Edição Comunitária

- A Wikipédia é um serviço baseado na tecnologia wiki, que permite que qualquer pessoa crie ou edite qualquer conteúdo aí publicado.



# Novas Ferramentas | Filtragem Comunitária

- Para responder à questão — **o que é importante?**, surgem novos serviços onde a comunidade colabora na identificação e seleção de informação.





# Novas Ferramentas | Filtragem Social

- Serão as recomendações dos amigos uma solução adequada para a seleção e organização de informação?



# Ideias Centrais

---

- A Internet introduziu uma rutura na produção, organização e distribuição de informação.
- O modelo pré-Internet (poucos produziam para muitos) está em extinção. O espaço público está a deslocar-se para a Internet, tornando-se acessível a qualquer indivíduo ou organização.
- As ferramentas anteriores não têm utilidade neste novo cenário disperso e fragmentado. Existe a necessidade (e oportunidade) de novas ferramentas.
  - **Como vamos gerir a quantidade de informação?**
  - **Como vamos encontrar o que nos importa?**
  - **Como vamos determinar a qualidade do que existe?**
- As tecnologias de informação têm um papel importante na resposta a estas questões.

# Referências

---

- **Computer Networks and Internets**  
Douglas E. Comer  
Prentice-Hall, 2004
- **Computers are Your Future**  
Catherine Laberta  
Pearson Education, 2010
- **Kevin Kelly: Predicting the next 5,000 days of the web**  
<http://youtu.be/yDYCf4ONh5M>