

Pesquisa no Passado

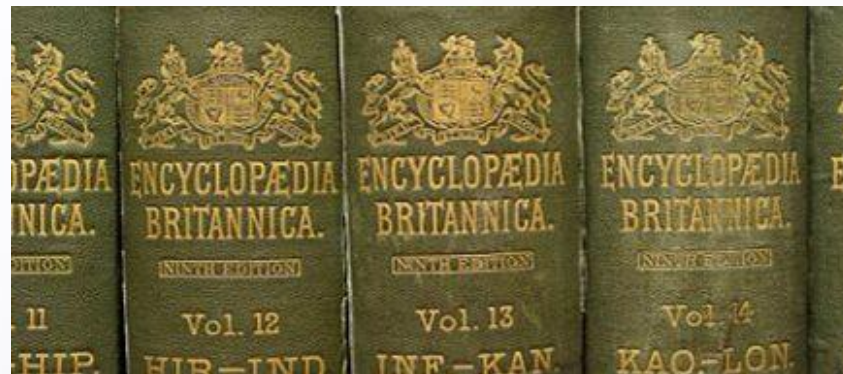
Miguel Costa

Engenheiro, Investigador, Aluno de Doutoramento

Faculdade de Letras da UL - 21 de Maio, 2012

Do Papel ao Digital

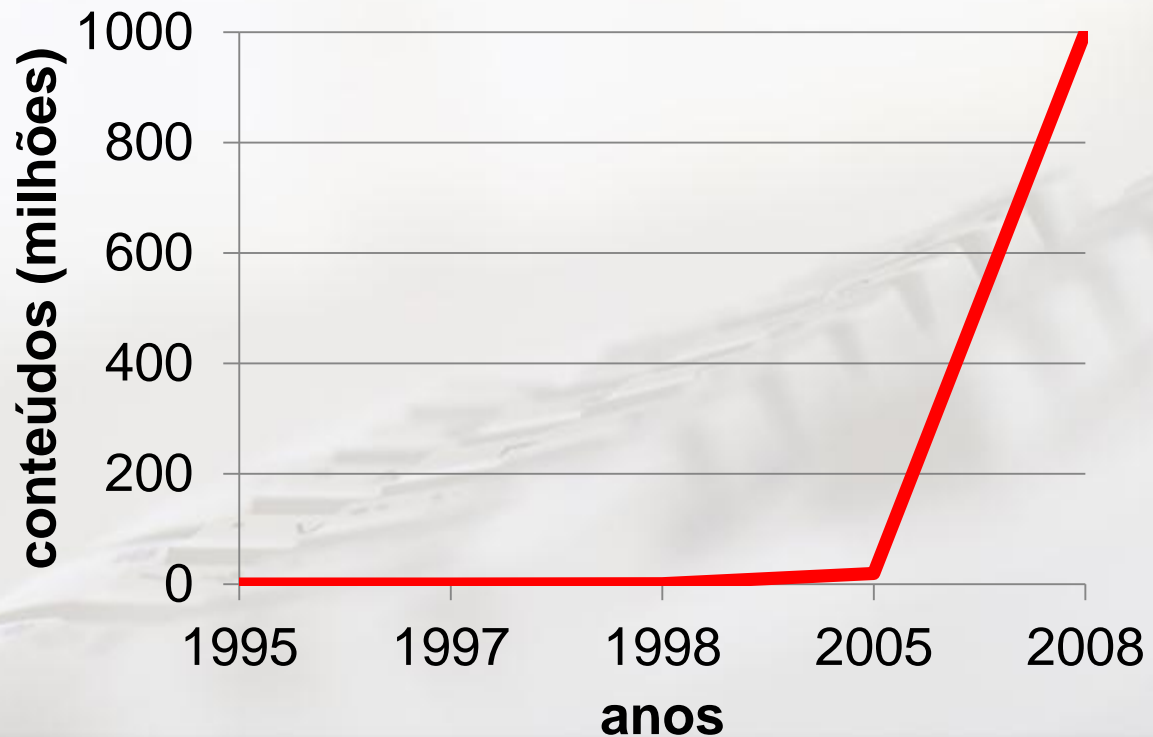
- Enciclopédia Britannica abandona edição impressa ao fim de 244 anos



- Livros → E-Books
- Diários → Blogs
- Álbuns de fotos → Flickr
- Cartas → Redes Sociais

Evolução da World Wide Web

- 1995: 11,4 milhões
- 1997: 200 milhões
- 1998: 800 milhões
- 2005: 19,2 mil milhões
- 2008: 1 bilião
- 2012: ???



Muitas Tecnologias Novas



Flash



Móvel



A Web é Efêmera

- 50 dias - 50% dos documentos são alterados
(Cho and Garcia-Molina. 2000)
- 1 ano - 80% dos documentos ficam inacessíveis
(Ntoulas, Cho and Olson. 2004)
- 27 meses - 13% citações para a Web desaparecem
(<http://webcitation.org/>. 2007)





404 - Página Não Encontrada



The page cannot be found

The page you are looking for might have been removed, had its name changed, or is temporarily unavailable.

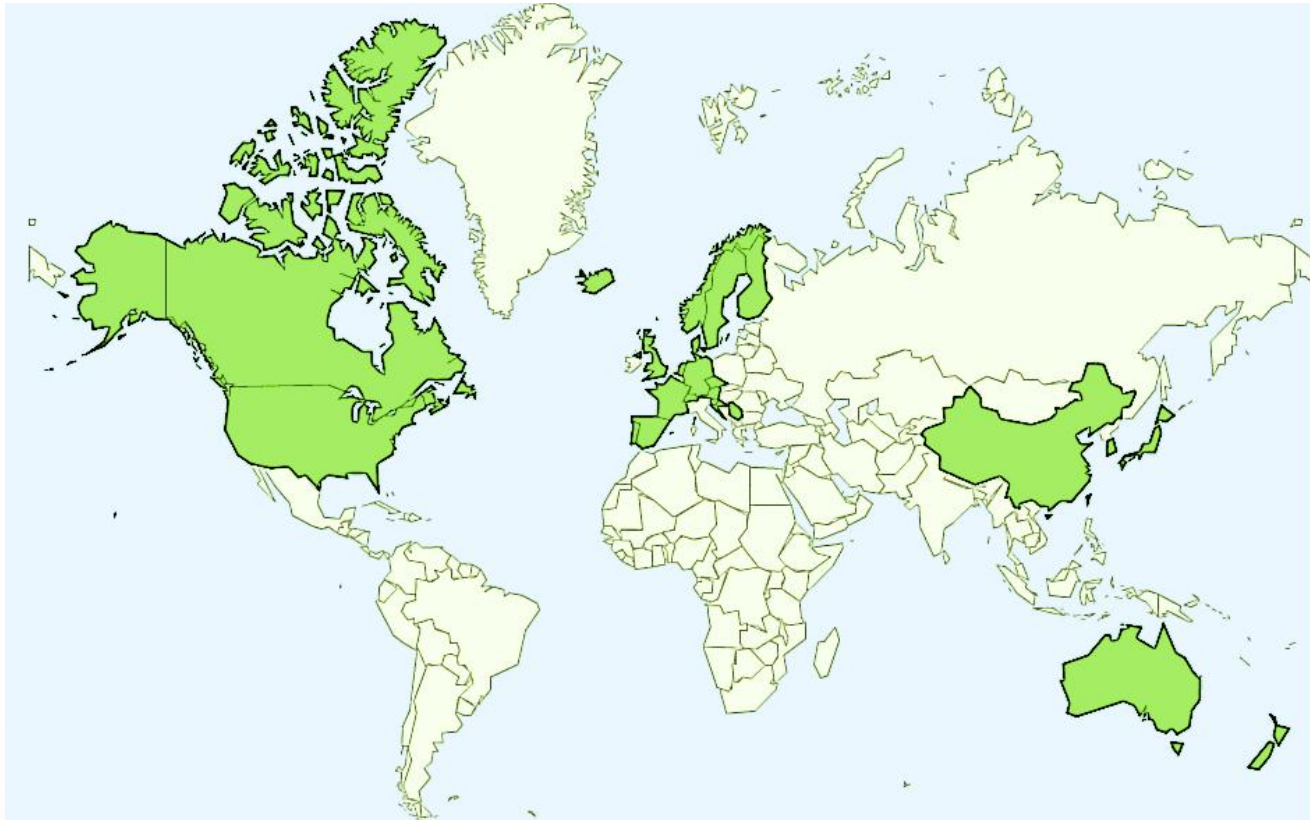
Please try the following:

- If you typed the page address in the Address bar, make sure that it is spelled correctly.
- Open the httpd.apache.org home page, and then look for links to the information you want.
- Click the  [Back](#) button to try another link.
- Click  [Search](#) to look for information on the Internet.

HTTP 404 - File not found
Internet Explorer

404
ERROR

Iniciativas de Arquivos da Web



- +77 iniciativas em 39 países
- +193 mil milhões de documentos desde 1996

Arquivo da Web Portuguesa



× Pesquisar no Arquivo

[Pesquisa avançada](#)

 **Pesquise e aceda a páginas do passado**
Veja ou reveja páginas que já desapareceram.

São mais de 130 milhões de páginas, arquivadas entre 1996 e 2010,
que estão ao seu dispor.

[Conheça o projeto](#)

- Disponível desde 2010: <http://arquivo.pt>
- Mil milhões de documentos
 - pesquisáveis por endereço (URL) e texto
 - entre 1996 e 2011

Pesquisa por Endereço (URL)



http://www.fl.ul.pt/



Pesquisar no Arquivo

entre:

01/01/1996



e:

31/12/2011



[Pesquisa avançada](#)

Versões da página web guardadas no arquivo

Foram gravadas no arquivo 248 versões da página <http://www.fl.ul.pt/> entre 1 Janeiro, 1996 e 18 Maio, 2012.

1999 1	2000 8	2001 15	2002 9	2003 11	2004 45	2005 68	2006 32	2007 36	2008 11	2009 6	2010 4	2011 2
25 Jan	1 Mar	21 Jan	30 Mai	7 Fev	9 Fev	4 Jan	4 Jan	3 Jan	1 Jan	23 Jun	5 Jun	22 Jan
	1 Mar	2 Fev	15 Jun	27 Fev	9 Fev	5 Jan	6 Jan	7 Jan	10 Jan	23 Jun	5 Jun	22 Jan
	3 Mar	17 Fev	26 Jul	10 Abr	25 Mar	6 Jan	12 Jan	12 Jan	19 Jan	25 Set	4 Ago	
	3 Mar	24 Fev	2 Ago	22 Abr	16 Abr	7 Jan	15 Jan	17 Jan	23 Jan	25 Set	4 Ago	



Estudantes

Acesso aos exames finais para alunos em regime normal de avaliação [\[mais informação\]](#) **NEW**

Atribuição de Bolsas Universidade de Lisboa / Fundação Amadeu Dias [\[mais informação\]](#) **NEW**

Cursos de 2º Ciclo de Estudos 2010/11 [\[mais informação\]](#) **NEW**

Despacho 6/2010: Determinação das datas do pagamento faseado de propinas referentes aos 2º e 3º ciclos da FLUL [\[mais informação\]](#)

Pagamento de Propinas na Tesouraria da FLUL [\[mais informação\]](#)

Esclarecimento anexo ao Regulamento Geral de Avaliação (20/10/2009) [\[mais informação\]](#)

Notícias

Cursos Intensivos de Verão (Julho) **NEW**
CLi - Centro de Línguas da FLUL [\[mais informação\]](#)

Curso Livre e Cursos de Verão **NEW**
Centro de Estudos Clássicos [\[mais informação\]](#)

ESC – English Summer Courses **NEW**
[\[mais informação\]](#)

Cursos de Verão 2010 **NEW**
Área de Ciências da Linguagem [\[mais informação\]](#)

Cursos de Português Língua Estrangeira **NEW**
Courses of Portuguese [\[mais informação\]](#)

Apresentação de novas Licenciaturas [\[mais informação\]](#)

Professores da Faculdade distinguidos com prémios [\[mais informação\]](#)

Agenda de Letras

Segue o [perfil da Faculdade](#) no **facebook**

[Colóquios / Congressos / Jornadas Científicas](#)

[Conferências / Seminários / Sessões](#)

[Exposições / Lançamento de Livros / Teatro / Concertos](#)

[Cursos Livres / Acções de Formação / Visitas de Estudo](#)

[Actos Académicos](#)

[Notícias](#)

[Bolsas e Programas de Investigação](#)

[Centenário da Faculdade de Letras](#)

[Ficha Técnica](#)



FACULDADE DE LETRAS DA UNIVERSIDADE DE LISBOA

Letras

Estudantes

Estudos Pós-Graduados 2004/2005 **NEW**
 Abertura de Candidaturas ao Mestrado em Geografia
[\[mais informação\]](#)

Propinas e Inscrições 2004/2005 [\[mais informação\]](#)

Actualização dos Horários 2004/2005 [\[mais informação\]](#)

Alunos SOCRATES / ERASMUS [\[mais informação\]](#)

Notícias

Curso Livre de Língua Gestual Portuguesa **NEW**
[\[mais informação\]](#)

CD delibera sobre valor das propinas [\[mais informação\]](#)

8ª Edição do Concurso de Ideias de Negócio da ANJE.
 Entrega de Candidaturas na UNIVA/FLUL [\[mais informação\]](#)

UNIVA tem novo horário de atendimento [\[mais informação\]](#)

[Plano de Actividades 2004](#)

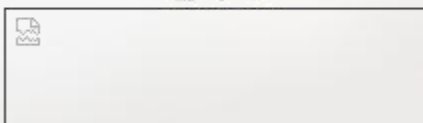
Eventos

Comemorações dos 150 anos do nascimento de Wenceslau Moraes [\[mais informação\]](#)

Colóquio: "Questions on the Linguistic Sign"
[\[mais informação\]](#)

I Congresso Virtual do Departamento de Literaturas Românicas: *Retórica* [\[mais informação\]](#)

2000



A visualização desta página em condições normais só é possível neste momento com o Internet Explorer 4x ou superior. Os utilizadores de qualquer versão do Netscape experimentarão falhas no html. Para fazer o download do Internet Explorer, clique no url www.ie4.com

Pesquisa Textual



faculdade de letras



Pesquisar no Arquivo

[Pesquisa avançada](#)

entre:



e:



Resultados 1 a 10 de 3.039.553

[Faculdade de Letras da Universidade de Lisboa](#)

25 Setembro, 2009 - [outras datas](#)

Faculdade de Letras da Universidade **de** Lisboa Exames **de** Conclusão **de** Licenciatura Prorrogação do ... , remunerações e avaliação dos trabalhadores da FLUL que exercem funções públicas A Agenda **de** Letras é o meio electrónico **de** divulgação das actividades culturais e científicas da **Faculdade de** Letras da ...

<http://www.fl.ul.pt/>

[Faculdade de Letras da Universidade de Lisboa](#)

20 Novembro, 2003 - [outras datas](#)

Faculdade de Letras da Universidade **de** Lisboa This page uses frames, but your browser doesn't support them. ...

<http://www.fl.ul.pt/>

[Faculdade de Letras » Faculdade de Letras » Universidade de ...](#)

3 Agosto, 2010 - [outras datas](#)

Faculdade de Letras » Faculdade de Letras » Universidade **de** Coimbra Universidade **de** Coimbra | www ... Antigos Estudantes Empresas Comunicação Social Visitantes e Agentes Turísticos **Faculdade de** Letras ... Profissionais Publicações Prémios Concursos Documentos para consulta Contactos UC **Faculdade de** Letras

...

<http://www.uc.pt/fluc>



Pesquisa Textual



faculdade de letras × Pesquisar no Arquivo

entre:  e:  [Pesquisa avançada](#)



entre:  e: 

[Faculdade de Letras da Universidade](#)

25 Setembro, 2009 - [outras datas](#)

Faculdade de Letras da Universidade de Lisboa Exames de Conclusão de Licenciatura Prorrogação do ... , remunerações e avaliação dos trabalhadores da FLUL que exercem funções públicas A Agenda de Letras é o meio electrónico de divulgação das actividades culturais e científicas da Faculdade de Letras da ...

<http://www.fl.ul.pt/>

[Faculdade de Letras da Universidade de Lisboa](#)

20 Novembro, 2003 - [outras datas](#)

Faculdade de Letras da Universidade de Lisboa This page uses frames, but your browser doesn't support them. ...

<http://www.fl.ul.pt/>

[Faculdade de Letras » Faculdade de Letras » Universidade de ...](#)

3 Agosto, 2010 - [outras datas](#)

Faculdade de Letras » Faculdade de Letras » Universidade de Coimbra Universidade de Coimbra | www ... Antigos Estudantes Empresas Comunicação Social Visitantes e Agentes Turísticos Faculdade de Letras ... Profissionais Publicações Prémios Concursos Documentos para consulta Contactos UC Faculdade de Letras

...

<http://www.uc.pt/fluc>

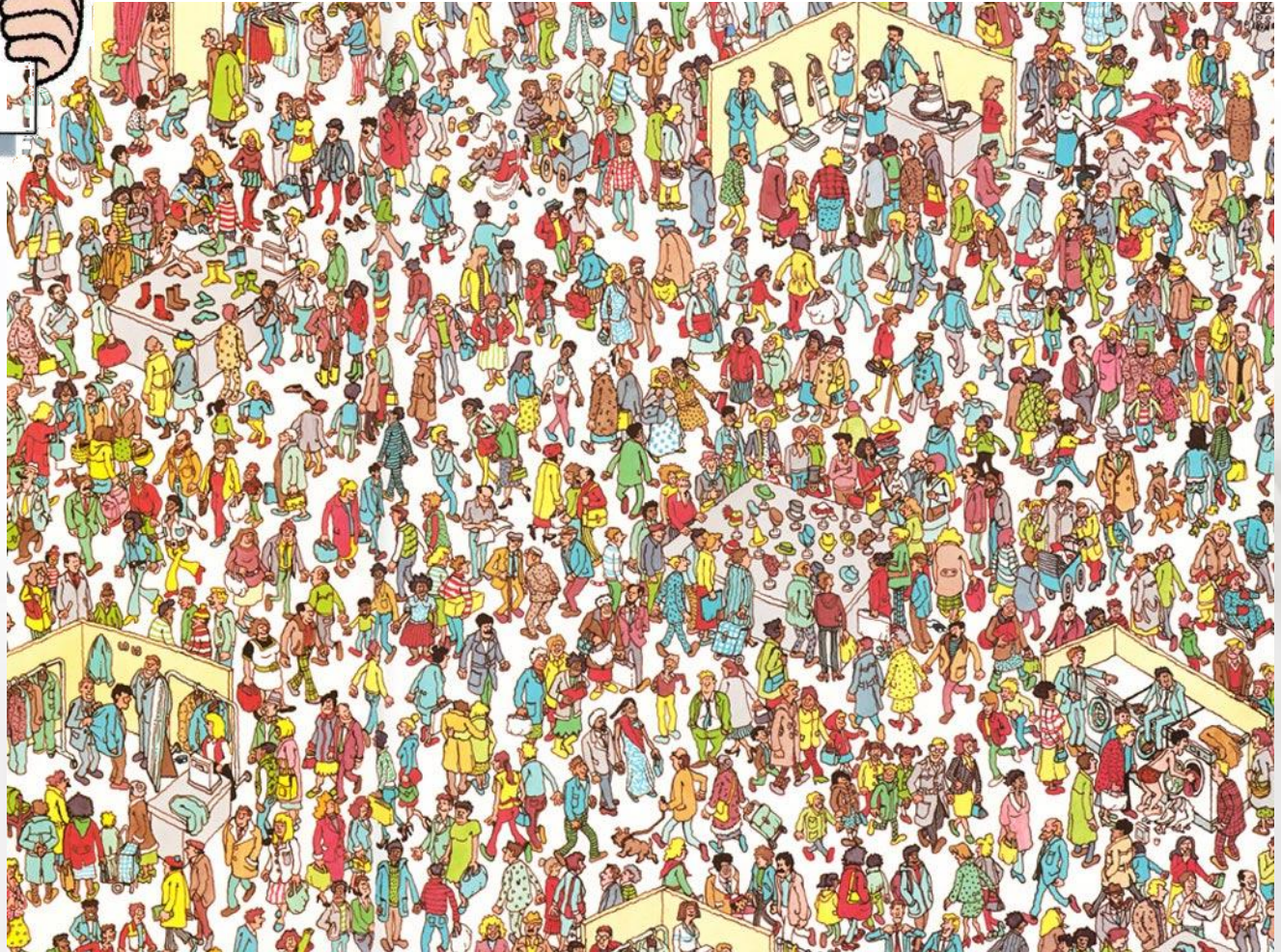
Complementar aos Motores de Busca



Casos de Uso

- Utilizador visita *Favorito* desaparecido
- Jornalista revisita notícia passada
- Webmaster recupera versão perdida do site
- Historiador analisa documentos digitais
- Web designer faz portfólio de sites antigos
- Professor recupera slides
- **Outros?**

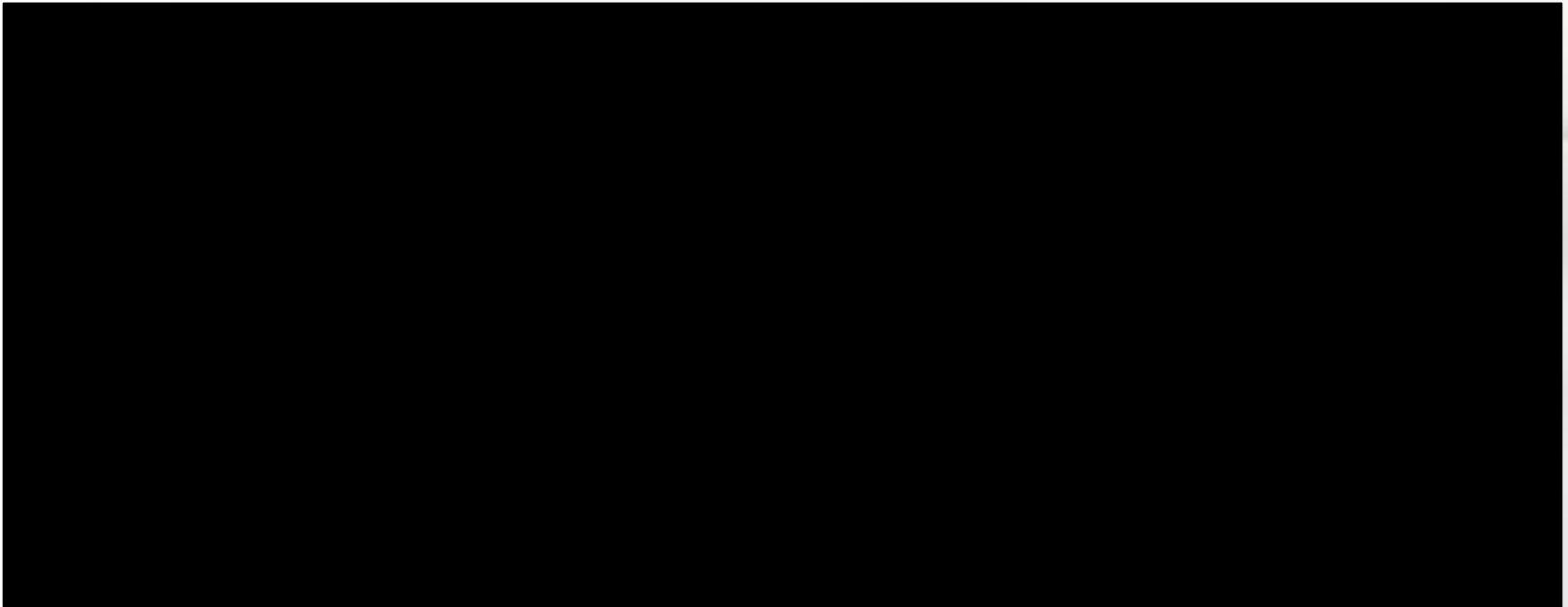
Onde está a Informação?



Como é feito a Pesquisa?

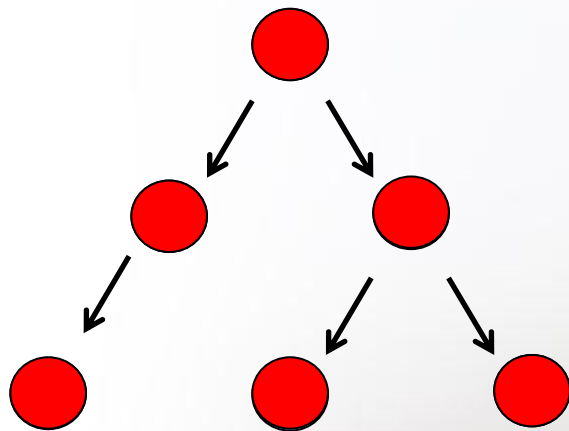


Pesquisa ↓ ↑ Resultados



Recolha de Dados

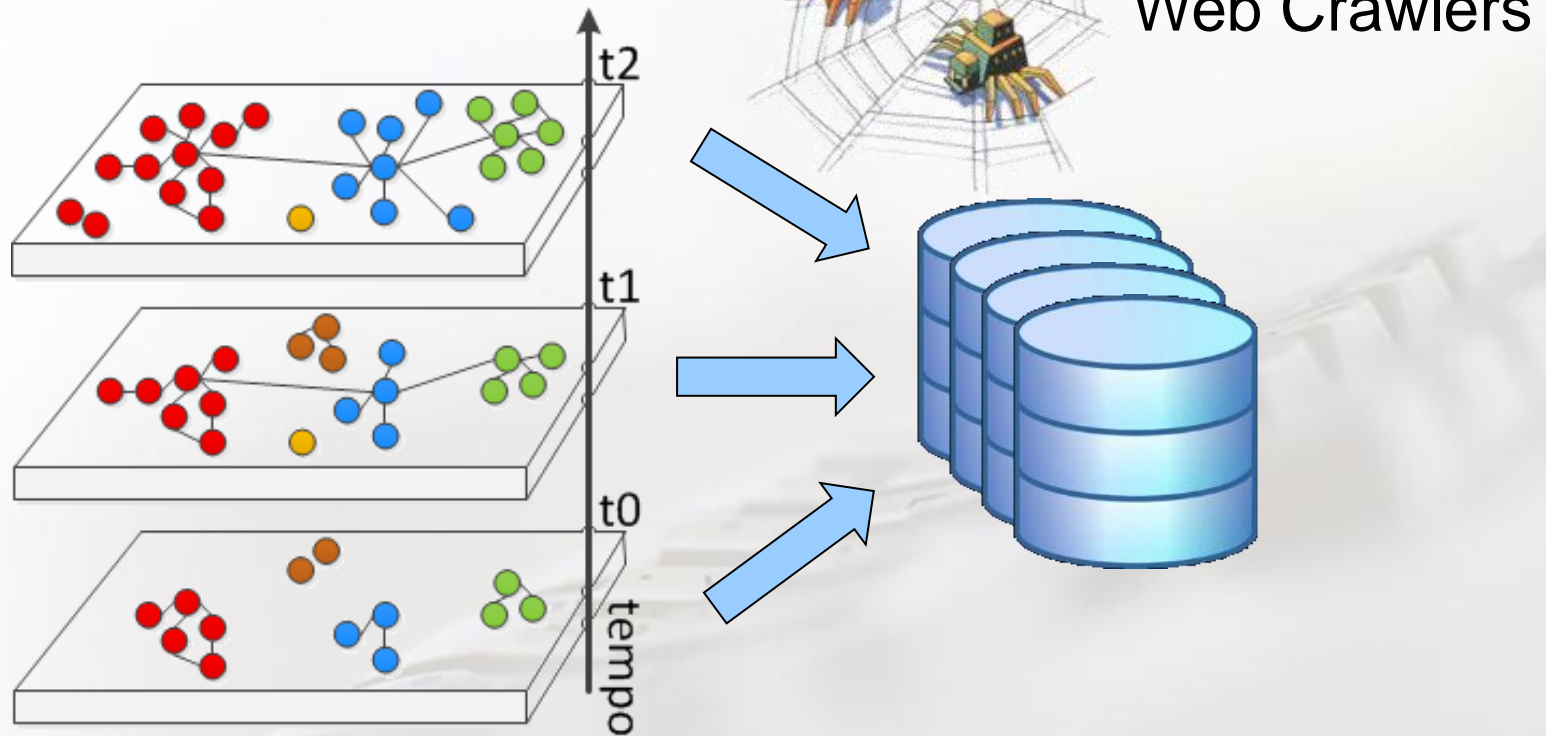
- **Automática** a partir de um conjunto de endereços
- Recursivamente são **seguidas ligações** e arquivados os conteúdos



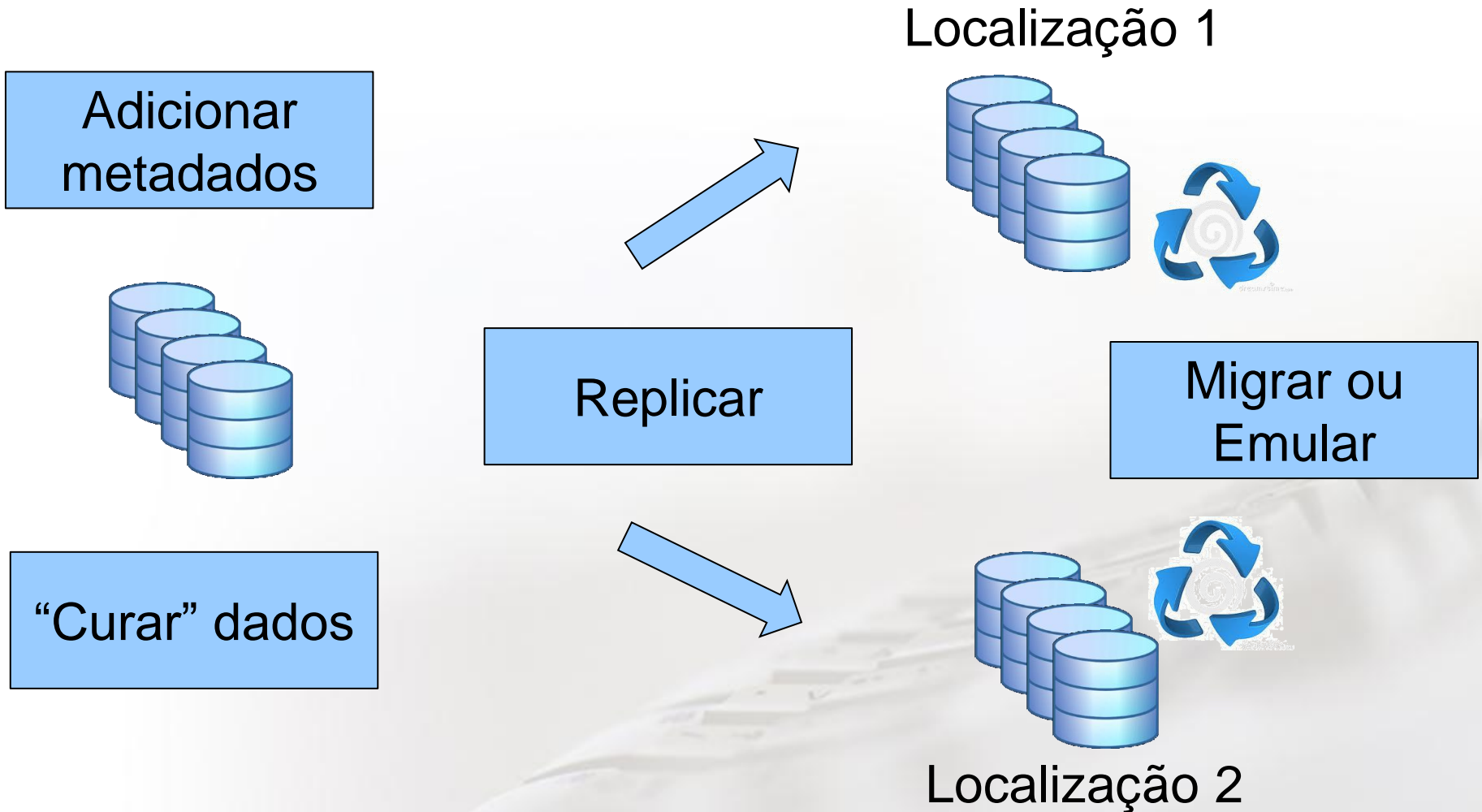
Web Crawlers

Recolha de Dados

- **Automática** a partir de um conjunto de endereços
- Iterativamente são **seguidas ligações** e arquivados os conteúdos



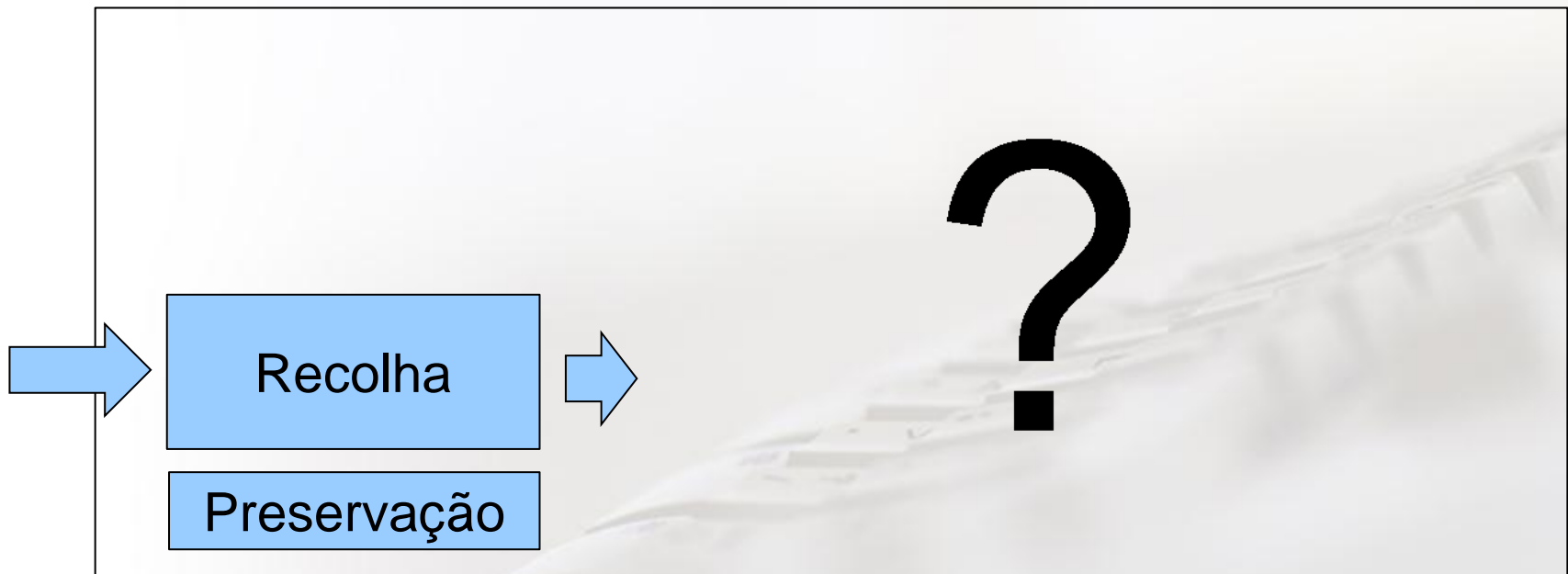
Preservação de Dados



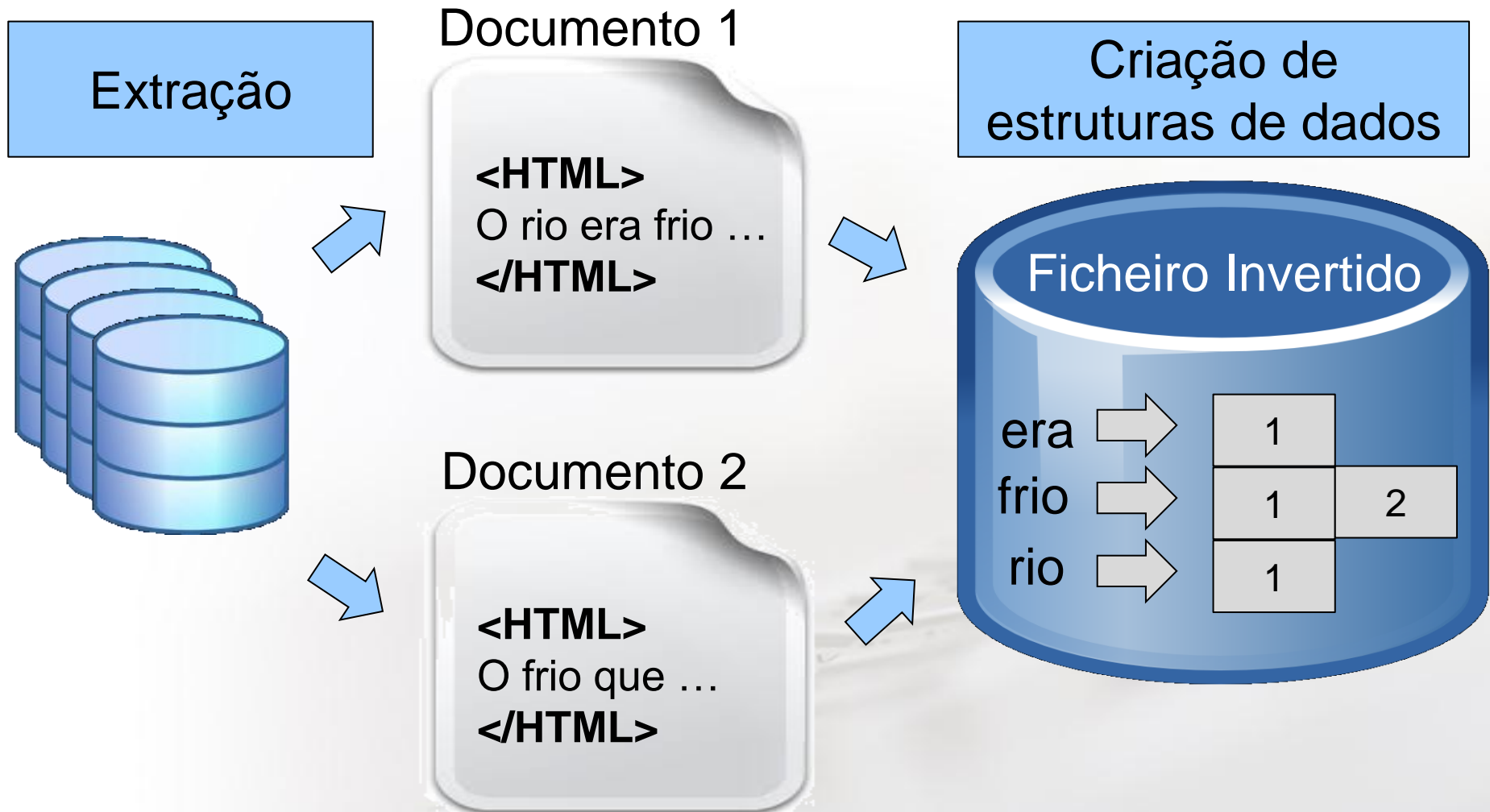
Como é feito a Pesquisa?



Pesquisa ↓ ↑ Resultados



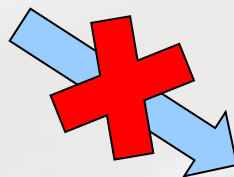
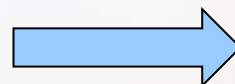
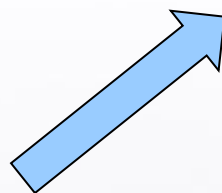
Indexação



Recuperação



rio frio @ [1996, 2000]



era	→	1		
frio	→	1	2	
rio	→	1		

1996

era	→	321	765	
frio	→	1041	1841	
rio	→	599	1841	

2000

era	→	2045	3012	3455
frio	→	2078	2090	
rio	→	4033		

2004

Ordenação



faculdade de letras × Pesquisar no Arquivo

entre: e:

[Pesquisa avançada](#)

Resultados 1 a 10 de 3.039.553

[Faculdade de Letras da Universidade de Lisboa](#)

25 Setembro, 2009 - [outras datas](#)

Faculdade de Letras da Universidade de Lisboa Exames de Conclusão de Licenciatura Prorrogação do ... , remunerações e avaliação dos trabalhadores da FLUL que exercem funções públicas A Agenda de Letras é o meio electrónico de divulgação das actividades culturais e científicas da Faculdade de Letras da ...

<http://www.fl.ul.pt/>

[Faculdade de Letras da Universidade de Lisboa](#)

20 Novembro, 2003 - [outras datas](#)

Faculdade de Letras da Universidade de Lisboa This page uses frames, but your browser doesn't support them. ...

<http://www.fl.ul.pt/>

[Faculdade de Letras » Faculdade de Letras » Universidade de ...](#)

3 Agosto, 2010 - [outras datas](#)

Faculdade de Letras » Faculdade de Letras » Universidade de Coimbra Universidade de Coimbra | www ... Antigos Estudantes Empresas Comunicação Social Visitantes e Agentes Turísticos Faculdade de Letras ... Profissionais Publicações Prémios Concursos Documentos para consulta Contactos UC Faculdade de Letras ...

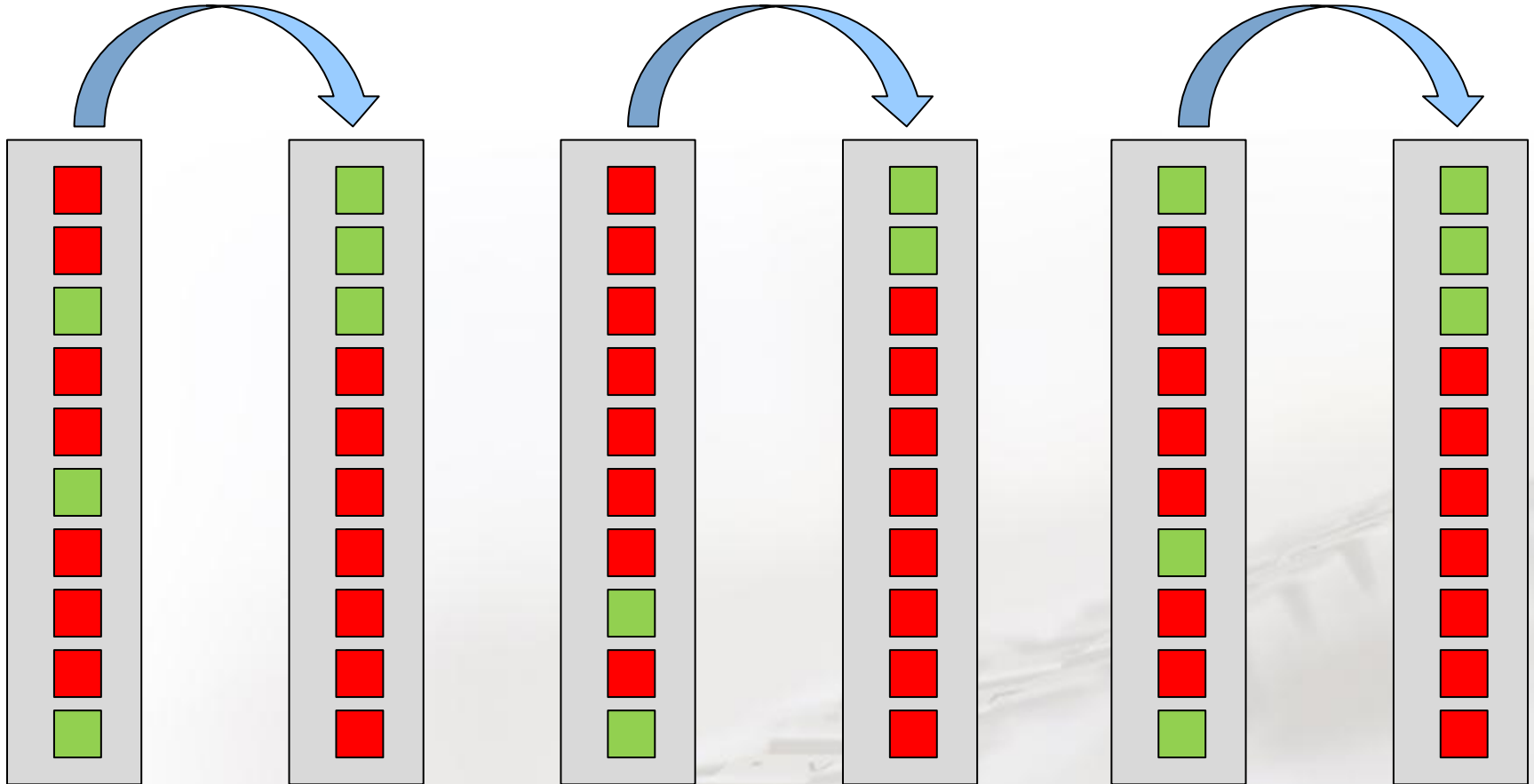
...

<http://www.uc.pt/fluc>

3.039.553

Objetivo

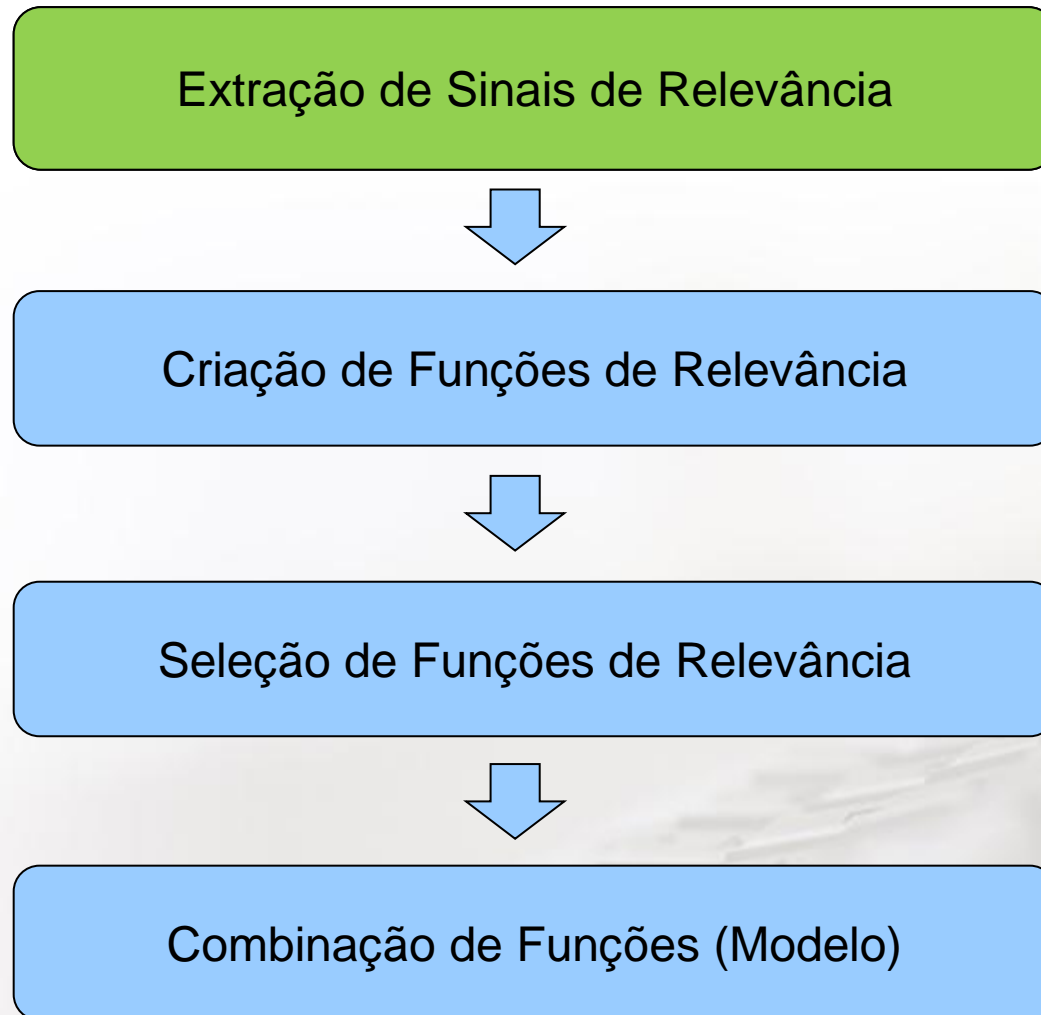
Maximizar Relevância



■ Relevante
■ Não Relevante

Para todas as pesquisas

Ordenação



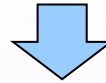
Sinais de Relevância

- Os termos da pesquisa estão:
 - quantas vezes no título?
 - próximo entre eles no texto?
- A página:
 - tem muitas versões diferentes?
 - tem muitos links a apontar para ela?
- Os utilizadores:
 - clicaram quantas vezes a página?
 - quanto tempo viram a página?



Ordenação

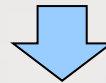
Extração de Sinais de Relevância



Criação de Funções de Relevância

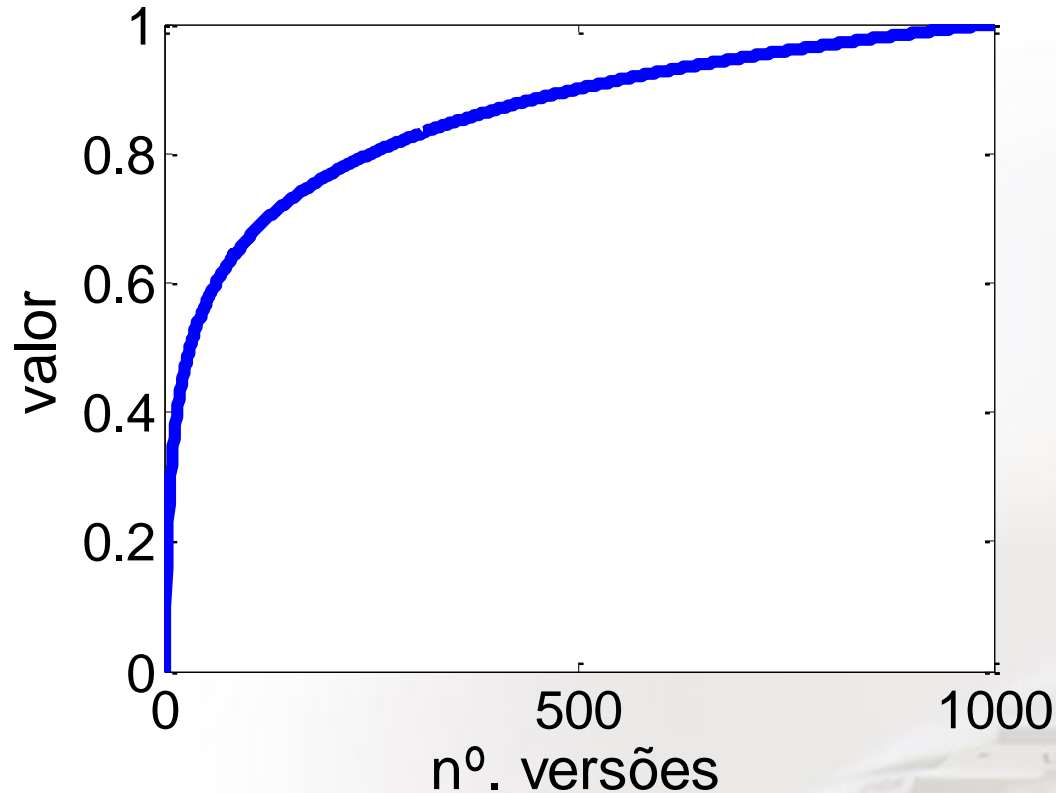


Seleção de Funções de Relevância



Combinação de Funções (Modelo)

Modelar os Sinais de Relevância



Exemplo de nº. de versões:

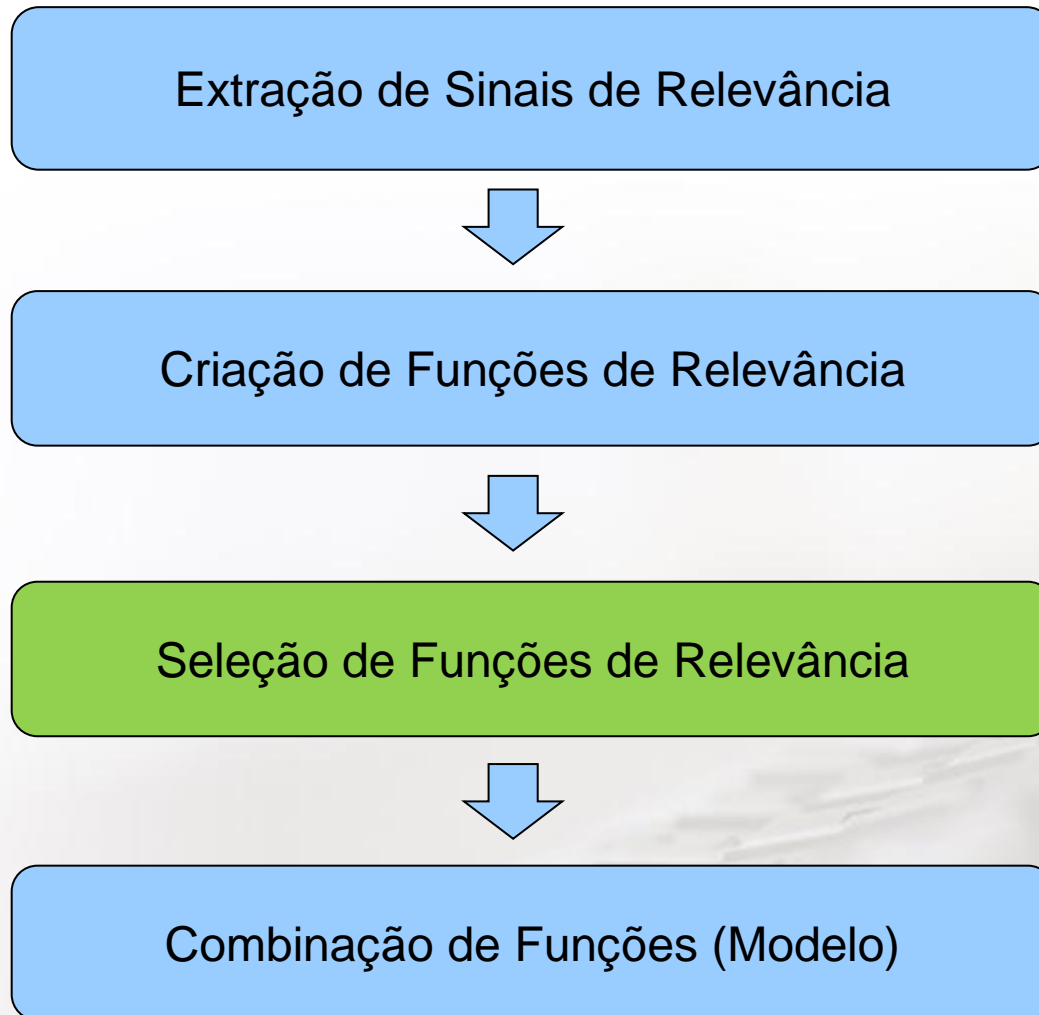
$$f(v_t^d) = \frac{\log_{10}(x)}{\log_{10}(y)} = \log_y(x)$$

Parâmetros:

$x = \text{nº. versões da página } d$

$y = \text{máximo nº. de versões}$

Ordenação

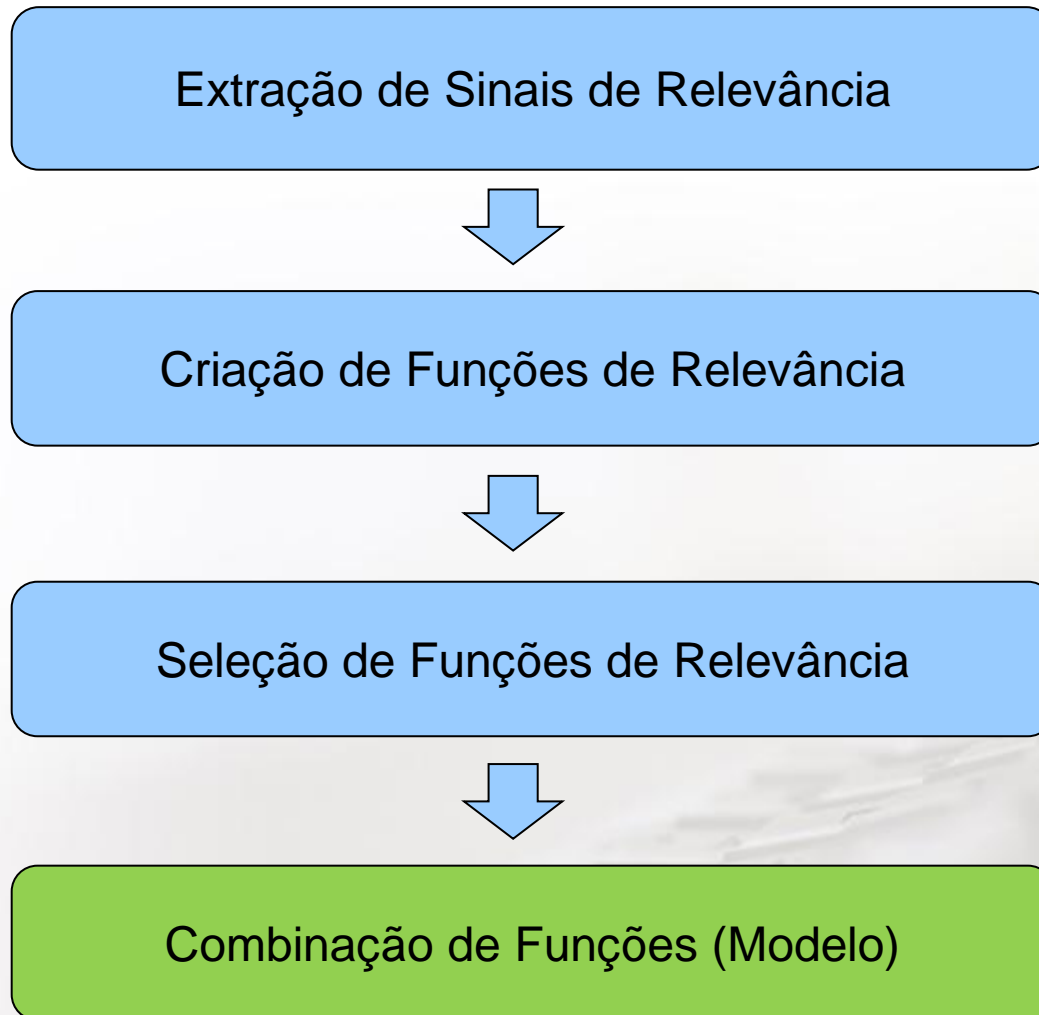


Selecionar Funções

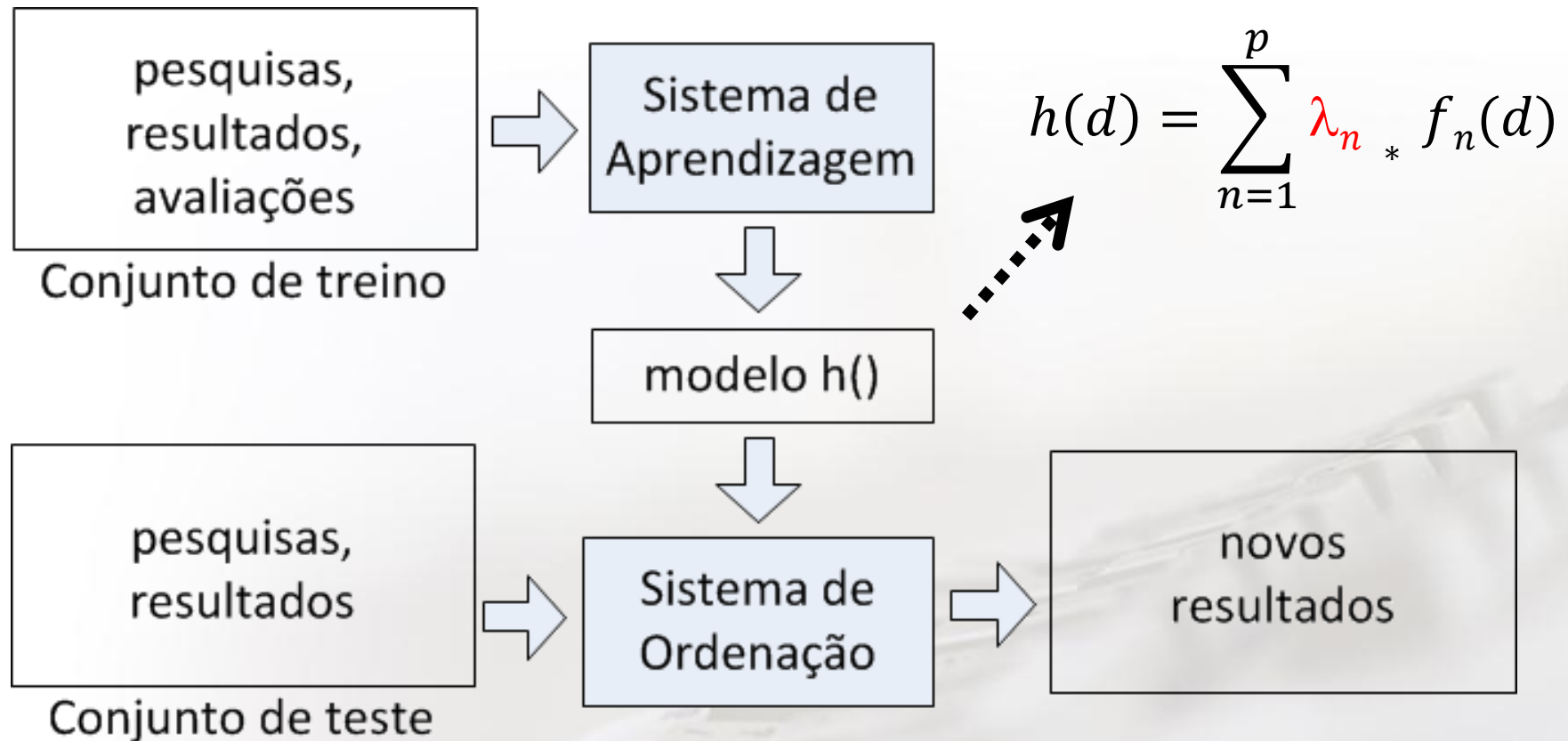
- Com bons resultados
- Não redundantes
- Computacionalmente leves



Ordenação



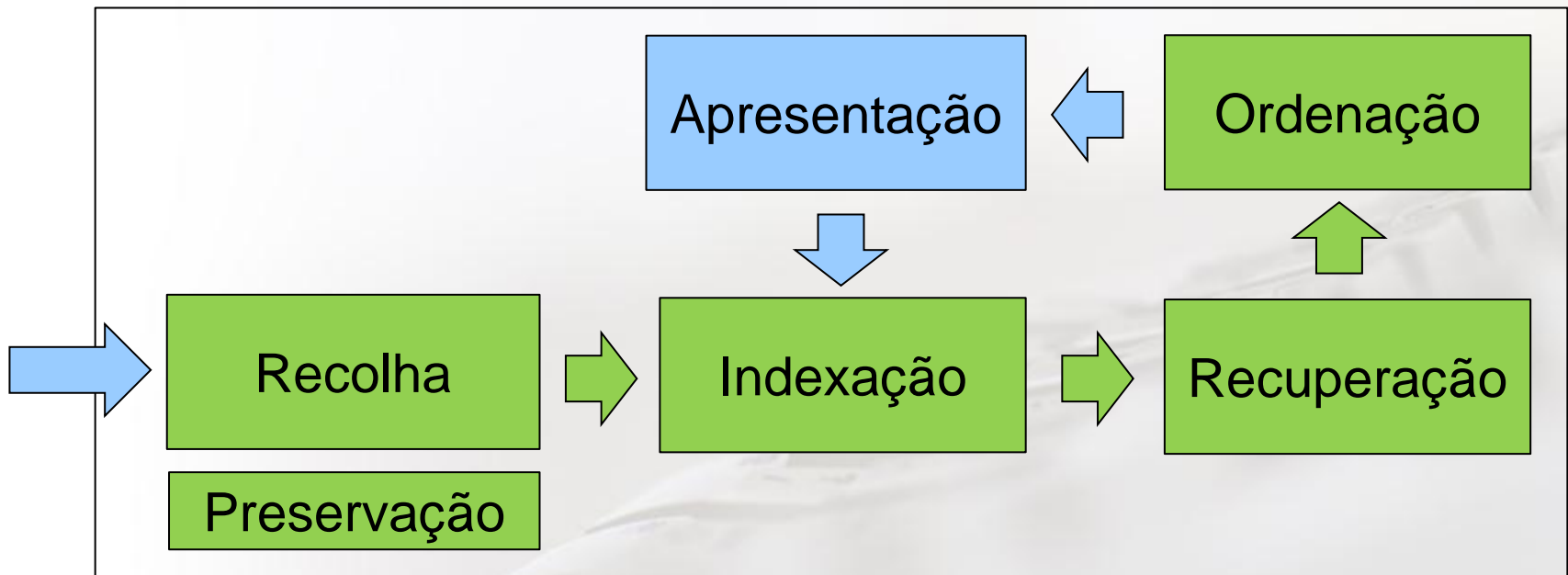
Combinar Funções



Como é feito a Pesquisa?



Pesquisa ↓ ↑ Resultados



Para saber Mais

- Ver: <http://sobre.arquivo.pt/>
 - publicações
 - vídeos
- Seguir notícias: <http://sobre.arquivo.pt/news>
 - Twitter
 - Facebook
 - LinkedIn
 - RSS feeds

- Seguir recomendações para autores
 - Não esconder ligações
 - Usar formatos adequados
- Sugerir sítios web portugueses para arquivo
- Fornecer conteúdos históricos
- **Divulgar!**

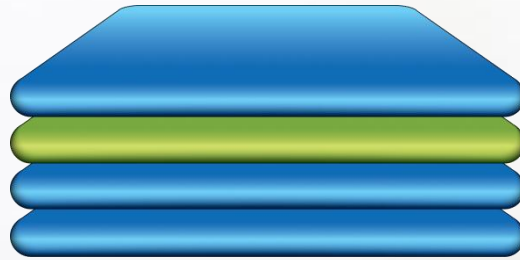
Desafios e Projetos

- Pesquisa de imagens
- Classificação automática de conteúdos
- Etiquetagem comunitária de conteúdos
- Reconhecimento de entidades mencionadas

- Para saber mais:

<http://www.arquivo.pt/propostas>

Obrigado



<http://arquivo.pt>

miguel.costa@fccn.pt

1996: SAPO



Servidor de Apontadores Portugueses

Pesquisa Global [Opções](#)

- **[Novidades](#)**
[Novos Links no SAPO](#), [Congressos](#), [Exposições](#), [Cursos de Formação](#), ...
- **[Ensino e Investigação](#)**
[Universidades](#), [Institutos](#), [Escolas](#), [Centros](#), ...
- **[Comunicação Social](#)**
[Jornais](#), [Rádios](#), [Televisão](#), [Revistas](#), ...
- **[Entretenimento](#)**
[Desportos](#), [Fotografia](#), [Passatempos](#), [Livros](#), [Fora de Casa](#), [Eduacional](#), ...
- **[Serviços de Informação](#)**
[Arquivos de Software](#), [Usenet News](#), [Talkers](#), [Archie](#), [Gopher](#), ...
- **[Comércio, Indústria e Serviços](#)**
[Serviços](#), [Informática](#), [Turísticas](#), [Financeiras](#), [Transportes](#), ...
- **[Páginas Pessoais](#)**
[Páginas pessoais de Portugueses](#), [Lista de Endereços de Correio Electrónico](#)
- **[Sociedade e Cultura](#)**
[Museus](#), [Bibliotecas](#), [Eventos](#), [Hospitais](#), [Religião](#), [Tributos](#), [Grupos e Associações](#), [Ministérios e Direcções Gerais](#), ...
- **[Regional](#)**
[Terras de Portugal](#), [Regiões de Turismo](#), [Açores](#), [Madeira](#), [Macau](#), [Timor](#), ...
- **[Computadores e Internet](#)**
[Documentação](#), [Web Designers](#), [Software](#), [Internet Service Providers](#), [Jogos](#), ...



Correio



Info



E-Mails



Roleta



Destaque



Pesquisa



Adicionar

[Centro de Informática](#) da [Universidade de Aveiro](#)

Internet URL- <http://sapo.ua.pt:80/>

1998: Abertura da EXPO 98

PRIMEIRA PÁGINA	REVISTA	OPINIÃO	POLÍTICA	SOCIEDADE
ECONOMIA	INTERNACIONAL	DESPORTO	CARTAZ	VIDAS

REVISTA

16/5/98

CAPA

Enfim, a Expo!

A cinco dias da abertura oficial, a Expo faz contas, revê estratégias e tapa os buracos. Foi para isso que serviu a antestreia. Por um dia, a festa esteve por conta do povo, em traje de passeio e a «ver os peixes».



À EXPO-98, que sábado passado viveu o seu dia de ensaio geral, já foram feitos todos os elogios e quase todos os avisos. É uma obra que enaltece a capacidade dos portugueses, toda a gente sabe, até porque nasceu do nada, ou melhor, foi levantada do lodo. Não há ali nada que seja a fingir.

Três dias antes da antestreia, o ministro dos Assuntos Parlamentares, António Costa (responsável pela «pasta» da Expo), avisava que «o lamaçal da insinuação não é o nosso habitat natural». Estava a falar de outra coisa mas, para o caso, é dito

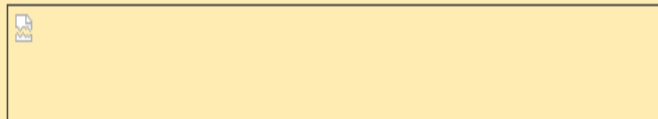


A Expo passou no teste, mas o exame é daqui a cinco dias

2001: Figo melhor jogador do mundo

O JOGO_{online}

Nº 300 / 17 Ter, 18 Dez 2001



Figo Melhor Jogador do Mundo

Reacções

Nuno Gomes:

"O Figo merece. Já o ano passado talvez merecesse ganhá-lo. Mas conquistou a Bola de Ouro; este ano é eleito como melhor jogador do Mundo, é um feito notável, que é motivo de orgulho para todos os portugueses. Os meus parabéns sinceros para o Figo, que fez mais uma grande época, provou que era o melhor entre os melhores."

Abel Xavier:

"É um motivo de grande orgulho. Todos os portugueses devem sentir-se satisfeitos e orgulhosos com este prémio. Estou muito contente, porque sou amigo dele há muitos anos. Figo recebe o prémio que talvez merecesse o ano passado. O Luís é uma grande pessoa, e em termos futebolísticos uma referência mundial. Para além disso é grande exemplo para todos os que vão começar e os que vão terminar as suas carreiras. Ele é o expoente máximo do futebol português."

Fernando Santos:



RELACIONADOS

O melhor do Mundo

Figo:
"Estou feliz por ter sido um português a ganhar"

Os melhores de sempre

2003: Portugal apoia a guerra no Iraque

[Página Inicial](#) [Primeira Vez](#) [Procurar](#) [Contacte-nos](#) [Ajuda](#) [Mapa](#) [Sobre o Portal](#)

[Imprimir](#) | [Recomendar](#)

[Modo Texto](#) | [English](#)

» [Primeiro-Ministro](#)

Declaração do Primeiro-Ministro sobre a crise iraquiana

S. Bento, 20 de Março de 2003

Acaba de ser desencadeada pelos Estados Unidos da América e por alguns dos seus aliados uma acção militar contra o regime iraquiano.

Este é um conflito que não desejámos. Tudo fizemos para o evitar, por meios políticos e diplomáticos.

Infelizmente não foi possível. A responsabilidade cabe exclusivamente ao regime iraquiano e à sua obstinação em não cumprir as resoluções impostas desde há doze anos pelas Nações Unidas.

Nesta hora difícil, Portugal reafirma o apoio aos seus aliados, com quem compartilha os valores da Liberdade e da Democracia, e faz votos de que esta seja uma acção tão rápida quanto possível e que cumpra todos os seus objectivos.

Espero e desejo que a vida dos inocentes possa ser poupada e que o conflito termine com o mínimo de sofrimento para todas as partes.

Como sempre dissemos, Portugal não envolverá neste conflito quaisquer forças militares. Cabe-nos, no âmbito político, honrar os compromissos para com os nossos aliados através

2004: Grécia vence Euro 2004

Jornal de Notícias

Director: Frederico M. Mendes
Director de Redacção: José Leite Pereira
Directores Adjuntos: Alfredo Leite e David Pontes
Subdirector: António José Teixeira

Domingo, 4 de Julho de 2004

Pesquisa JN

Edição do Dia ▶ Últimas Notícias

Diário de Notícias | TSF Online

Arquivo | Dossier | Secção do Leitor

Edição do Dia

Envie este artigo para alguém | Salvar para ler depois | Imprimir este artigo

Em Foco

Sociedade

Política

Economia

Mundo

Cultura

Etcetera

Desporto

Televisão

Opinião

Tema da Semana

Sénior

Última Página

Grécia vence Euro 2004

III ▶ Selecção Nacional lutou até ao fim mas perdeu por 1-0

Portugal perdeu a final do Euro 2004 por 1-0 com a Grécia. O esquema defensivo bem montado da equipa grega levou a melhor sobre as fintas de Cristiano Ronaldo e companhia.

Golo grego surgiu aos 57 minutos, por Charisteas no seguimento de um pinta pé de canto. A partir daí a equipa portuguesa atacou ainda com mais vigor mas o resultado permaneceu inalterado até ao final.

■ Incêndios continuam a atacar Portugal

O que faz a FCCN? www.fccn.pt

Gestão da rede de Investigação e Ensino Nacional

Serviços de Rede



DNS



Segurança & Serviços à Comunidade



Serviços Avançados

