jornadas FCCN

Pesquisa de texto com Apache Solr

mais rápida, moderna e escalável

Vasco Rato

FCCN serviços digitais fct | fct Fundação para a Ciência e a Tecnologia | U.PORTO | U.PORTO FEP FACULDADE DE ECONOMIA UNIVERSIDADE DO PORTO

Pesquisa de texto no Arquivo.pt

- Implementada em 2010
- Primeiro arquivo da web a ter pesquisa de texto aberta

Pesquisa de texto no Arquivo.pt

Pesquisa de texto no Arquivo.pt

Menu | arquivo.pt | fct | Opções

faculdade de economia do porto | Pesquisar

2005 4 Nov — 2006 5 Mai

Páginas | Imagens | Narrativa | Pesquisa avançada

Cerca de 13.529 resultados desde **2005** até **2006**

www.fep.up.pt

FEP - **Faculdade de Economia do Porto**

24 Novembro 2005

FEP - **Faculdade de Economia do Porto** ...

www.infoforum.pt/web/pesquisa/resultados_mestrados.asp?

INFOFORUM - Informação

18 Dezembro 2005

Decisão UNIVERSIDADE **PORTO FACULDADE ECONOMIA PORTO** Mestrado/Curso **de** Especialização Análise ... AVEIRO Mestrado/Curso **de** Especialização Actividade Física Adaptada UNIVERSIDADE **PORTO FACULDADE** ... UNIVERSIDADE **PORTO FACULDADE** CIÊNCIAS DESPORTO E EDUCAÇÃO FÍSICA **PORTO** Mestrado/Curso **de** ...

www.cm-sjm.pt/files/30/3050.pdf

[PDF] Bolsas **de** EStudo.xls

19 Fevereiro 2006

Economia Faculdade Economia Porto (1) "... ter-se verificado o aproveitamento escolar **do** bolseiro ...
Faculdade Letras Univ **Porto** Não Renovado (1) 3 Daniela Filipa Borges **de** Pinho 4 143,07 Enfermagem Esc Catarina Gonçalves Pinho Candeias 4 147,66 12 17 Medicina **Faculdade** Medicina **Porto** 10 Bibiana Patricia

arquivo.pt | fct

Universidade do Porto — FEP Faculdade de Economia

Portuguese Finance Network — 6-8 July | FEP

Você está em: Início

Menu Principal
- Notícias
- Órgãos de Gestão
- Grupos
- Serviços
- Cursos
- I & D
- Cooperação
- Pessoal
- Alunos
- Pesquisa

Autenticação
Utilizador:
Senha:
Validar

Mapa do Campus

Faculdade de Economia da Universidade do Porto

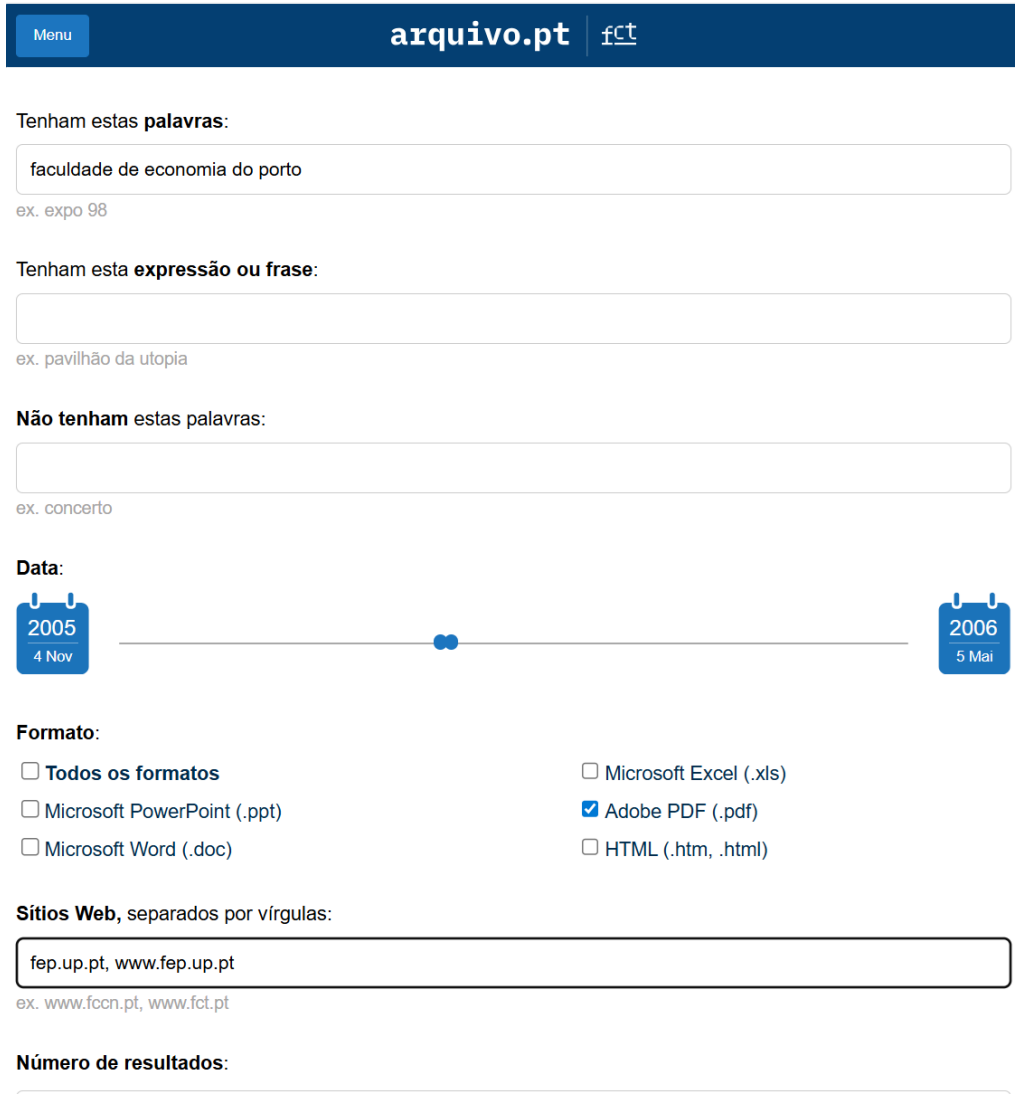
- Apresentação
- Serviço Informática
- Biblioteca
- ECTS Guide
- CONFERÊNCIAS

Últimas Notícias

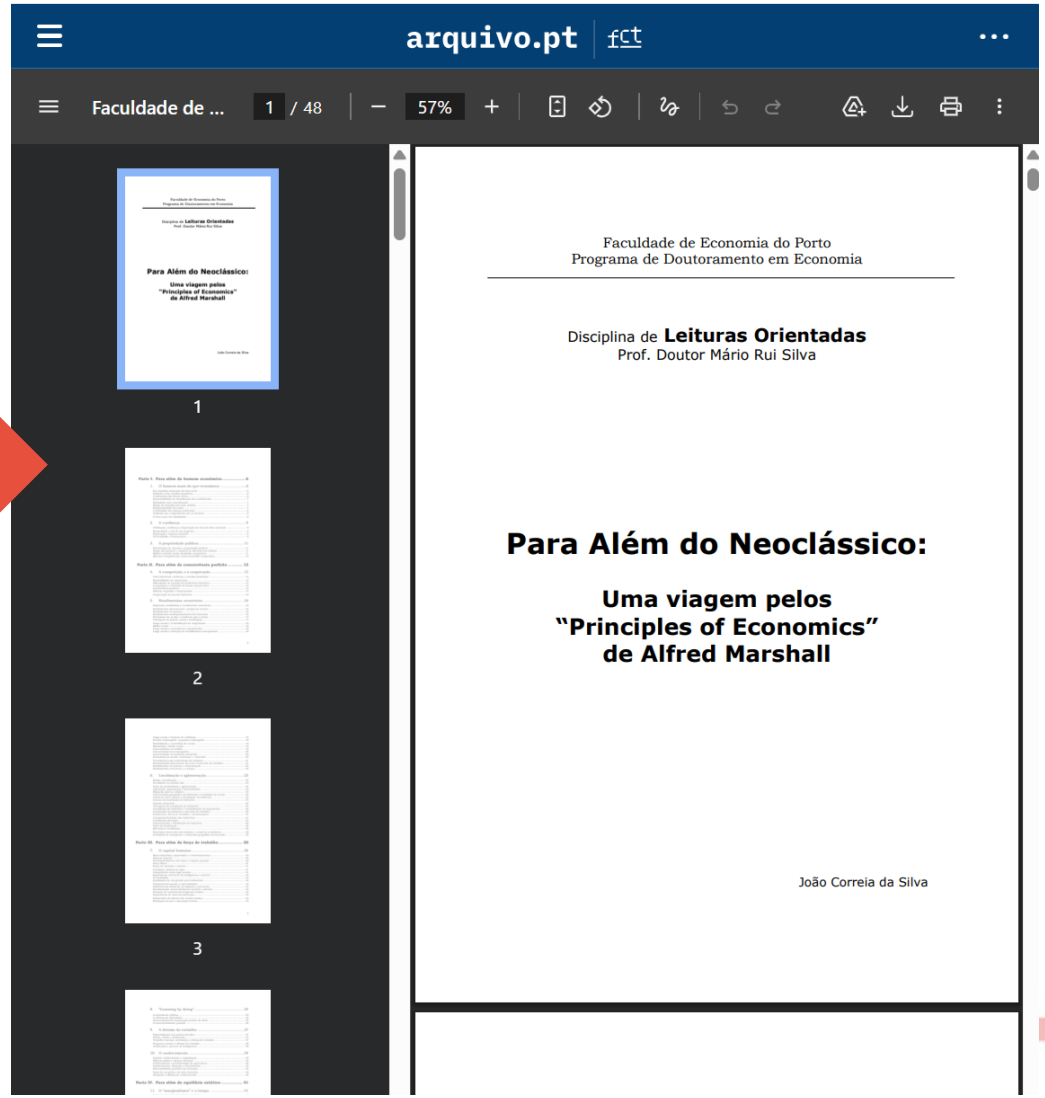
- Seminários de Finanças Cempre/Cete
  Liquidity and Market Efficiency . T. Chordia (Emory University, EUA), Richard Roll (UCLA) and Avanidhar Subrahmanyam (UCLA). Quinta-feira, 24 Novembro, 17h30, Sala Conselho. Abstract: The... [Ler toda a notícia].
- Defesa de tese do Mestrado em Análise de Dados e Sistemas de Apoio à Decisão
  Defesa de tese a ter lugar no dia 24 de Novembro de 2005 , pelas 11:00 , no Auditório 260, da candidata Licenciada Maria Fernanda Nogueira Gomes , tendo como título "SMAGeB: Sistema Multi-Agente... [Ler toda a notícia].
- Desafios e Inquietudes na Odisseia da Gestão de Recursos Humanos
  Seminário com Susana Vaz, no dia 25 de Novembro às 11h30m, no auditório 260 da Faculdade de Economia da Universidade do Porto. Susana Vaz é Directora de Recursos Humanos da Cifial, SA., com... [Ler toda a notícia].
- Concurso de ideias - Verão 2006
  Até ao final de Novembro, encontra-se aberto o Concurso de Ideias para actividades a desenvolver no âmbito dos Programas de Verão da Universidade Júnior. Todos os docentes e investigadores da U.P. são... [Ler toda a notícia].
- AJUDAS TÉCNICAS A ESTUDANTES COM DEFICIÊNCIA - UNIVERSIDADE DO PORTO
  Encontra-se aberto concurso para atribuição de ajudas técnicas aos estudantes com deficiência da UP. O concurso decorrerá entre 17 de Novembro a 5 de Dezembro. Entrega de candidaturas e... [Ler toda a notícia].

Opções
- Mapa de Páginas
- Seminários
- Conferências
- Working Papers
- Discussion Papers
- A FEP na Imprensa
- NotíciasFEP
- Calendário Escolar
- Calendário de Avaliação
- Horário de início e fim de aulas
- Webmail
- Rede Wireless e-U
- Consulta detalhada das Impressões
- Bolsa de Emprego
- Webforos
- Apontadores

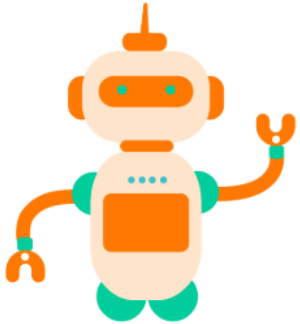
Pesquisa de texto no Arquivo.pt



Pesquisa de texto no Arquivo.pt

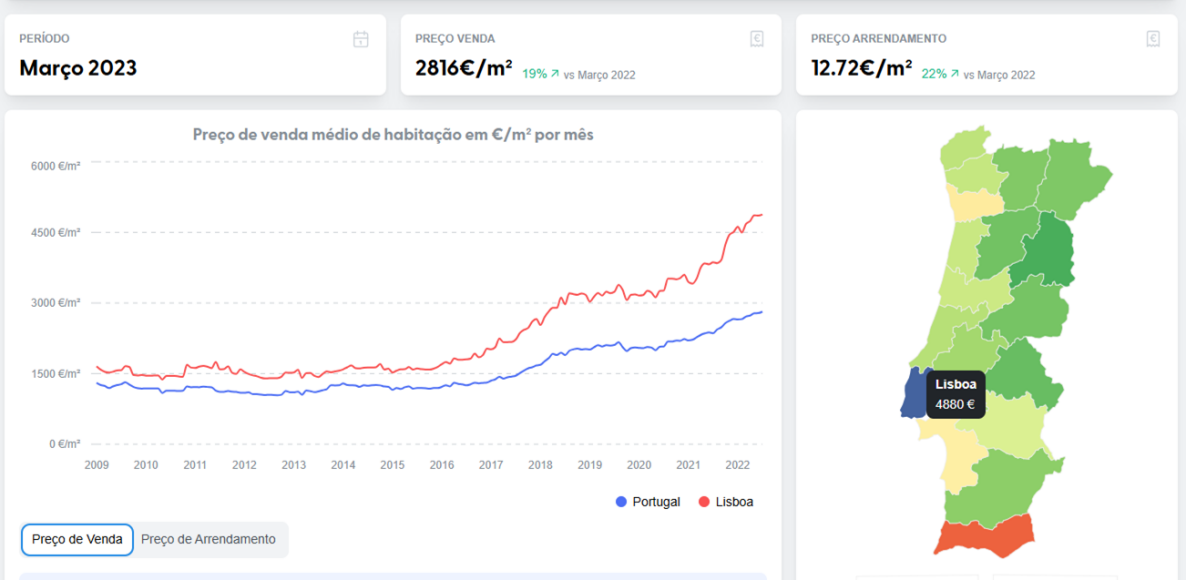


Pesquisa de texto no Arquivo.pt



API

2º lugar do prémio arquivo.pt em 2024



habitacao.net

arquivo.pt/api

Pesquisa de texto no Arquivo.pt

- Baseado em NutchWAX

- O projecto original já não existe…

Pesquisa de texto no Arquivo.pt

- Mas ainda está acessível no Arquivo.pt! ;)

Problemas do nutchWAX

- Requer manter outro software em versões estáticas para se manter compatível
- Projecto não é mantido pela comunidade - não queremos estar “orgulhosamente sós”
- Difícil de fazer alterações quando há problemas
- Etc…

O novo sistema - Solr

- Desde 2024 que temos vindo a refazer a pesquisa de texto em Apache Solr

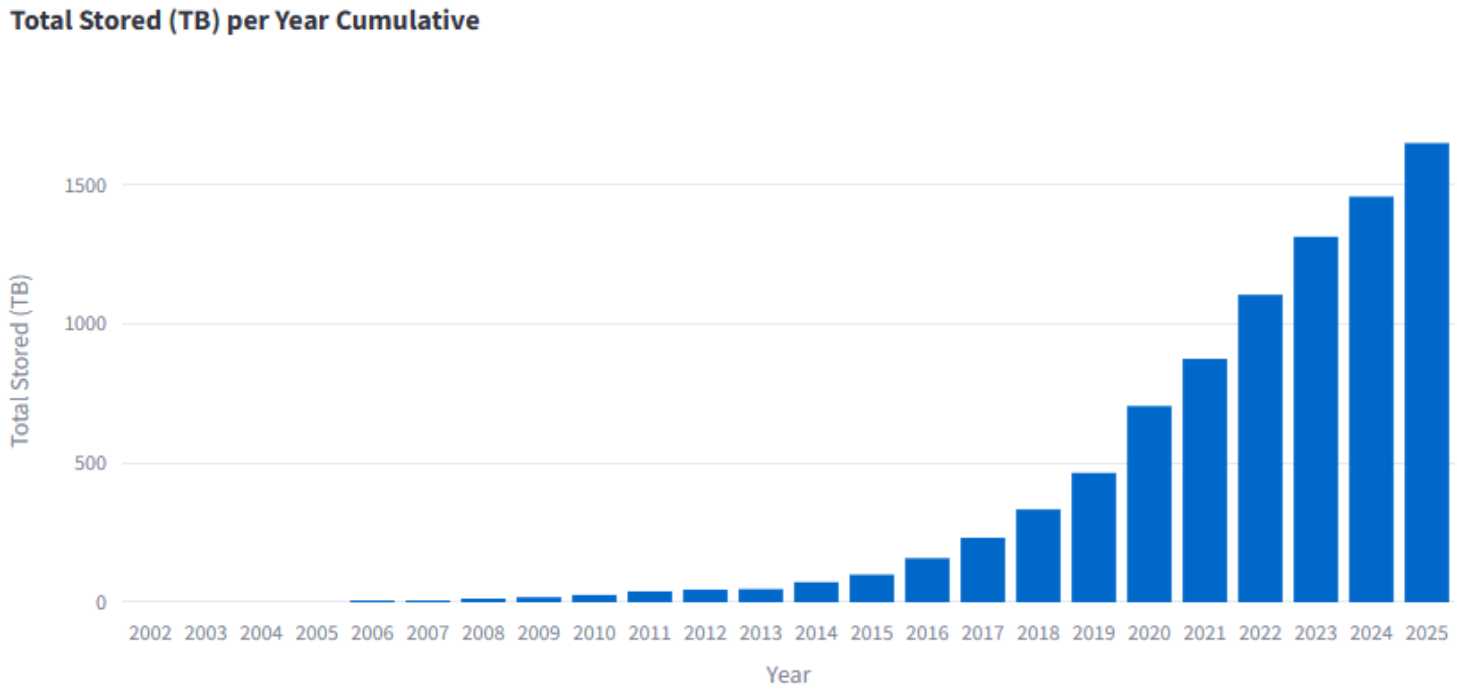
Porquê Solr?

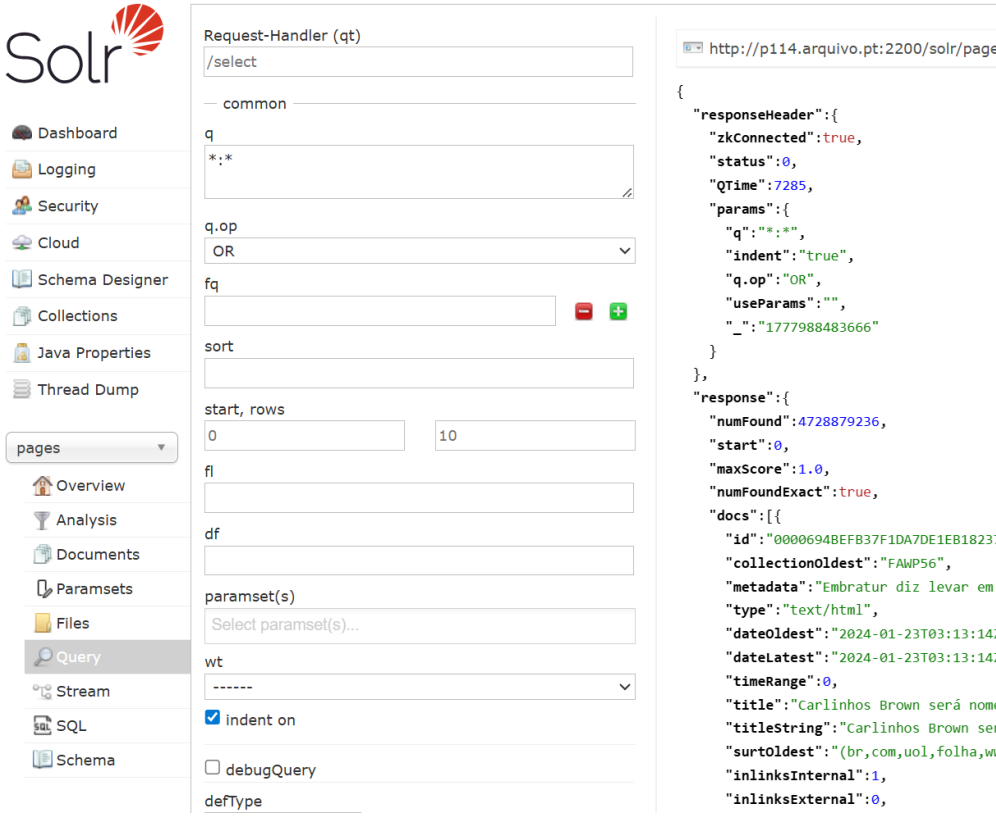
- Projecto open-source ativamente mantido pela comunidade
- Outros arquivos da web também têm soluções para pesquisa de texto baseadas em Solr
- O nosso sistema de pesquisa de imagens já é baseado em Solr
- Permite pesquisas mais complexas
- É mais escalável

Solr

Os desafios

- Adaptar outros componentes como a API para usar o novo sistema de pesquisa
- Reindexar mais de 1PB de dados…



Para quando?

- Adaptar os outros componentes ✅
- Reindexação ✅
- Protótipo ✅
- Substituição do sistema antigo - em breve!