



amália

LLM da Língua Portuguesa de Portugal

Assistente Multimodal Automático de Linguagem com IA

Estratégia de Inovação em IA



Draghi report, 2024

“Europe must profoundly refocus its collective efforts on **closing the innovation gap with the US and China.**”



European AI gigafactories, 2025

“AI will improve our healthcare, spur our research and innovation and boost our competitiveness. We want AI to be a force for good and for growth

***[European AI gigafactories], akin to a CERN for AI, will enable all our scientists and companies – not just the biggest - to develop the most advanced very large models needed to make Europe an AI continent.**”*

https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en

IA Generativa na Europa e em Portugal

— ou —

Como deve Portugal posicionar-se dentro da
Europa / afirmar-se na Europa como líder em
IA?



Diz o poema de Camões onde Leonor vai descalça à fonte



Diz o poema de Camões onde Leonor vai descalça à fonte



O poema que mencionas é um soneto muito conhecido de Luís de Camões, intitulado "**Descalça vai para a fonte Leonor**". Eis o texto completo:



Diz o poema de Camões onde Leonor vai descalça à fonte



O poema que mencionas é um soneto muito conhecido de Luís de Camões, intitulado **"Descalça vai para a fonte Leonor"**. Eis o texto completo:

Descalça vai para a fonte Leonor

Descalça vai para a fonte Leonor
Pelas veredas de estevas e de flores;
Saem-lhe ao rosto seus cabelos loiros,
Que amarelos são mais que o ouro.

Leva na cabeça o cântaro vazio
E vai cantando, alegre e despreocupada,
A harmonia da sua voz ao vento,
Que encanta o rústico povo.



Diz o poema de Camões onde Leonor vai descalça à fonte



O poema que mencionas é um soneto muito conhecido de Luís de Camões, intitulado **"Descalça vai para a fonte Leonor"**. Eis o texto completo:

Descalça vai para a fonte Leonor

Descalça vai para a fonte Leonor
Pelas veredas de estevas e de flores;
Saem-lhe ao rosto seus cabelos loiros,
Que amarelos são mais que o ouro.

Leva na cabeça o cântaro vazio
E vai cantando, alegre e despreocupada,
A harmonia da sua voz ao vento,
Que encanta o rústico povo.





Diz o poema de Camões onde Leonor vai descalça à fonte



O poema que mencionas é um soneto muito conhecido de Luís de Camões, intitulado **"Descalça vai para a fonte Leonor"**. Eis o texto completo:

Descalça vai para a fonte Leonor

Descalça vai para a fonte Leonor
Pelas veredas de estevas e de flores;
Saem-lhe ao rosto seus cabelos loiros,
Que amarelos são mais que o ouro.

Leva na cabeça o cântaro vazio
E vai cantando, alegre e despreocupada,
A harmonia da sua voz ao vento,
Que encanta o rústico povo.



Descalça vai para a fonte
Lianor pela verdura;
Vai fermosa, e não segura.

Leva na cabeça o pote,
O testo nas mãos de prata,
Cinta de fina escarlata,
Sainho de chamelete;
Traz a vasquinha de cote,
Mais branca que a neve pura.
Vai fermosa e não segura.

Descobre a touca a garganta,
Cabelos de ouro entrançado
Fita de cor de encarnado,
Tão linda que o mundo espanta.
Chove nela graça tanta,
Que dá graça à fermosura.
Vai fermosa e não segura.



Diz o poema de Camões onde Leonor vai descalça à fonte



O poema que mencionas é um soneto muito conhecido de Luís de Camões, intitulado **"Descalça vai para a fonte Leonor"**. Eis o texto completo:

Descalça vai para a fonte Leonor

Descalça vai para a fonte Leonor
Pelas veredas de estevas e de flores;
Saem-lhe ao rosto seus cabelos loiros,
Que amarelos são mais que o ouro.

Leva na cabeça o cântaro vazio
E vai cantando, alegre e despreocupada,
A harmonia da sua voz ao vento,
Que encanta o rústico povo.



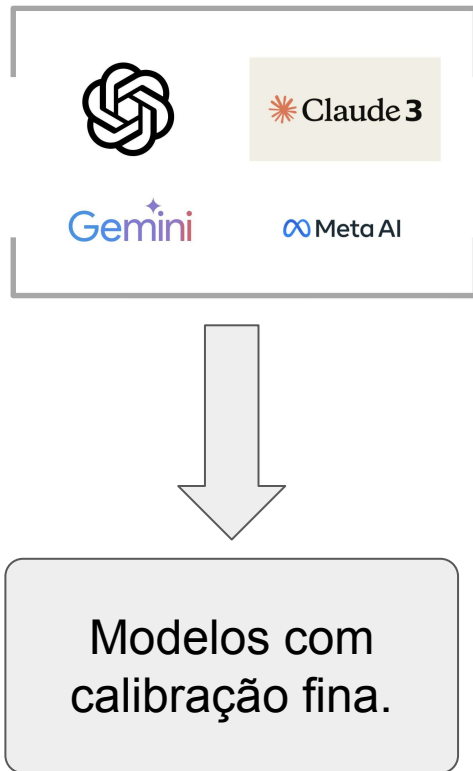
Descalça vai para a fonte
Lianor pela verdura;
Vai fermosa, e não segura.

Leva na cabeça o pote,
O testo nas mãos de prata,
Cinta de fina escarlata,
Sainho de chamelete;
Traz a vasquinha de cote,
Mais branca que a neve pura.
Vai fermosa e não segura.

Descobre a touca a garganta,
Cabelos de ouro entrançado
Fita de cor de encarnado,
Tão linda que o mundo espanta.
Chove nela graça tanta,
Que dá graça à fermosura.
Vai fermosa e não segura.

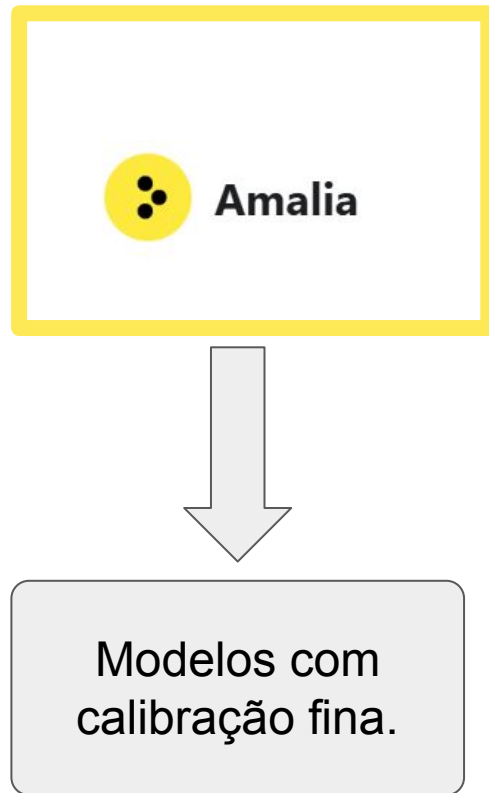
Porque precisamos de um LLM português?

- Modelos como o **ChatGPT**, **Claude**, ou **Gemini** são focados na língua e cultura anglo-saxónicas:
 - Apresentam pior desempenho e lacunas culturais quando usados noutras línguas
 - Têm dificuldade em distinguir variantes do português (de Portugal e do Brasil)
- Metodologias de baixo custo passam por calibrações finas de modelos pré-treinados.
 - Aumentam a dependência estratégica.
 - Transferem o saber e dados para terceiros.



Porque precisamos de um LLM português?

- A dependência de IA externa (dos EUA e da China) é uma vulnerabilidade e uma oportunidade perdida para desenvolver a nossa capacidade em IA.
- É importante assegurar a soberania nacional e europeia no desenvolvimento de IA — no âmbito da [Estratégia Nacional de IA](#).
- É necessário desenvolver LLMs **abertos** e **multilíngues**:
 - que todos possam usar e melhorar;
 - que representem as línguas e os valores europeus.



A oportunidade e o momento para criar um LLM PT-PT

- “... demoraria demasiado tempo e exigiria recursos computacionais excessivos para treinar um LLM”



Deucalion + Marenostrium 5

- “... não temos dados suficientes”



- “... não temos o conhecimento necessário”



O caminho até o AMALIA



GlorIA 1B, 2023

Primeiro modelo PT-PT

Pré-treino parcial PT-PT

Dados do arquivo.pt



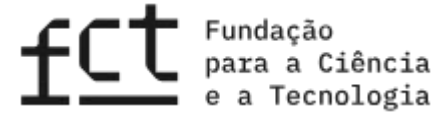
EuroLLM 9B, 2024

Modelo para todas as
línguas da EU

Pré-treino de raíz

Experiência no EuroHPC

Consórcio AMALIA



Colaborações





amália

LLM da Língua Portuguesa de Portugal

Assistente Multimodal Automático de Linguagem com IA

AMALIA



Controlo sobre a língua portuguesa

Permitindo que empresas nacionais, Administração Pública e demais entidades se expressem corretamente nas várias variantes do português.



Soberania sobre os dados de cidadãos

Possibilitando aplicar os mais recentes avanços em IA a domínios como a Administração Pública, garantindo que dados confidenciais não saiam de território nacional.



Preservar a representatividade cultural de Portugal

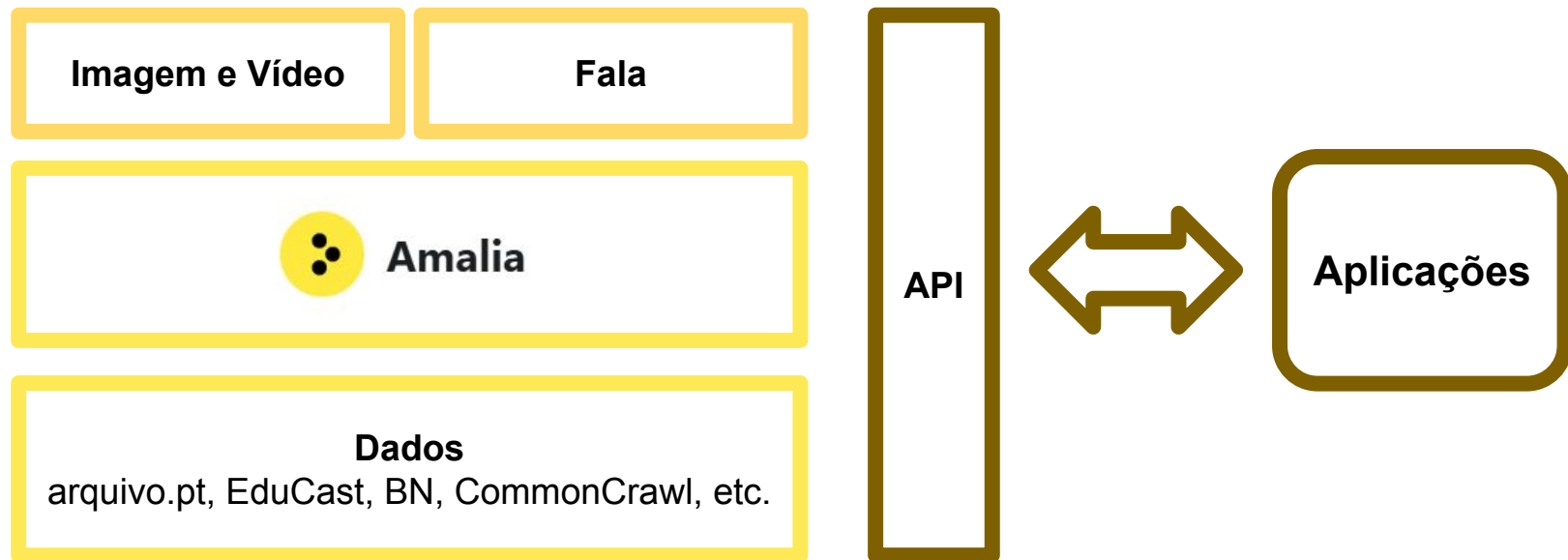
Capturando e preservando expressões idiomáticas, gírias e referências culturais próprias de Portugal.



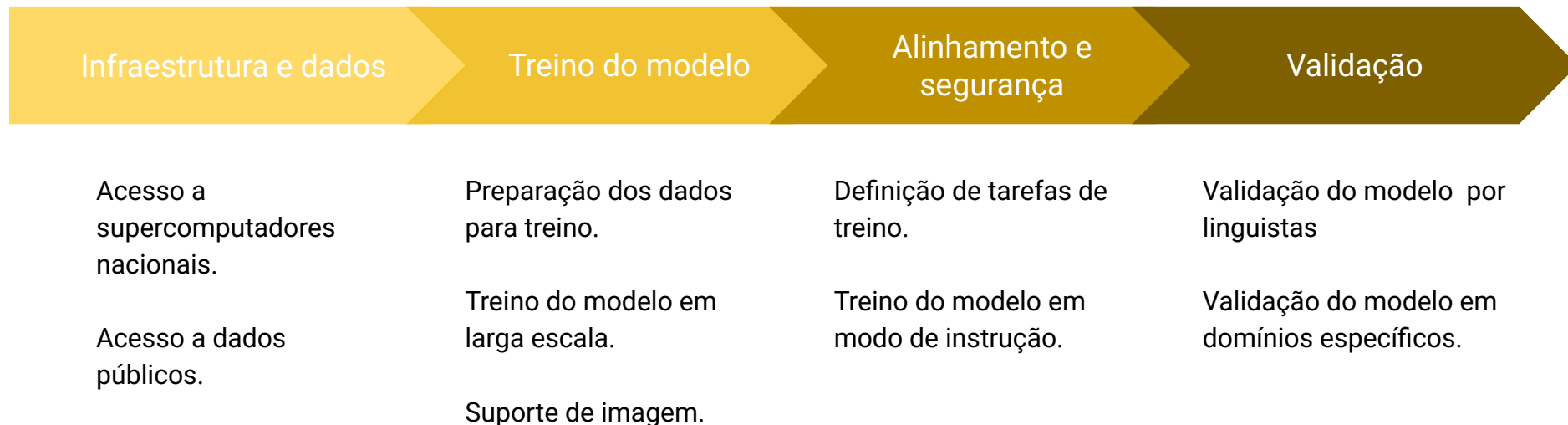
Promoção da investigação e inovação em IA

Estimulando a investigação e inovação colaborativa em IA, juntando vários centros de investigação nacionais.

Arquitectura do AMALIA



Processo de desenvolvimento do AMALIA



Validação do modelo

Avaliação automática e manual por linguistas:

- Avaliação do modelo geral
- Tarefas de texto: PR, RAG, diálogo, tradução
- Tarefas multimodais: PR visual, descrição visual, pesquisa multimodal

A validação e a extensão do modelo será ainda realizada em domínios específicos pelos parceiros do consórcio.



Domínios Específicos

Educação

**Museus e
Cultura**

Ciência

Media



amália



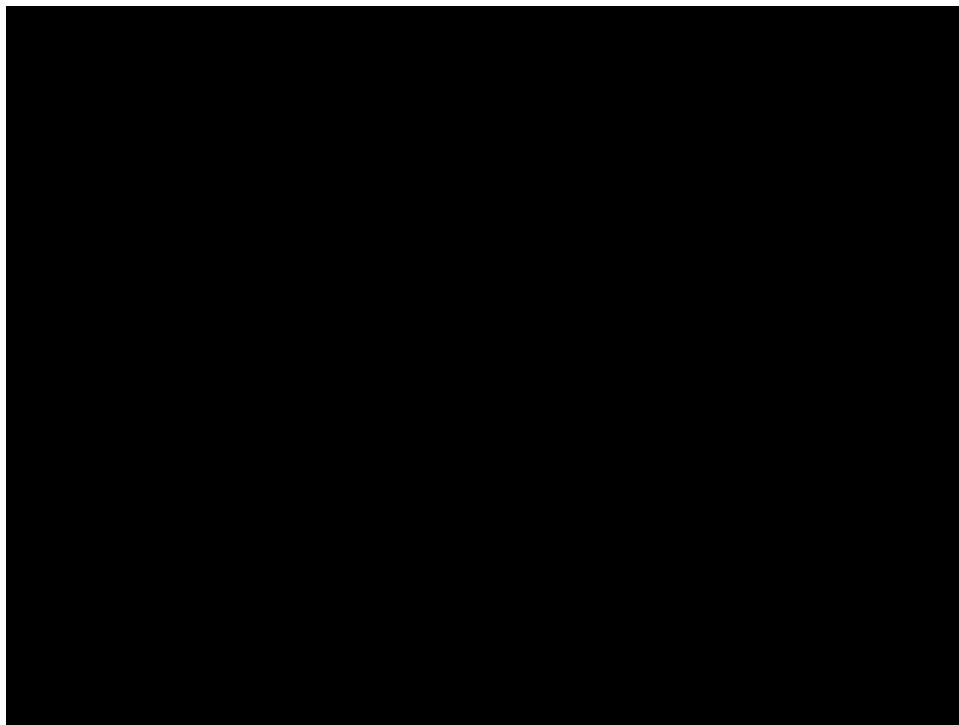
Universidade do Minho
uminho.pt



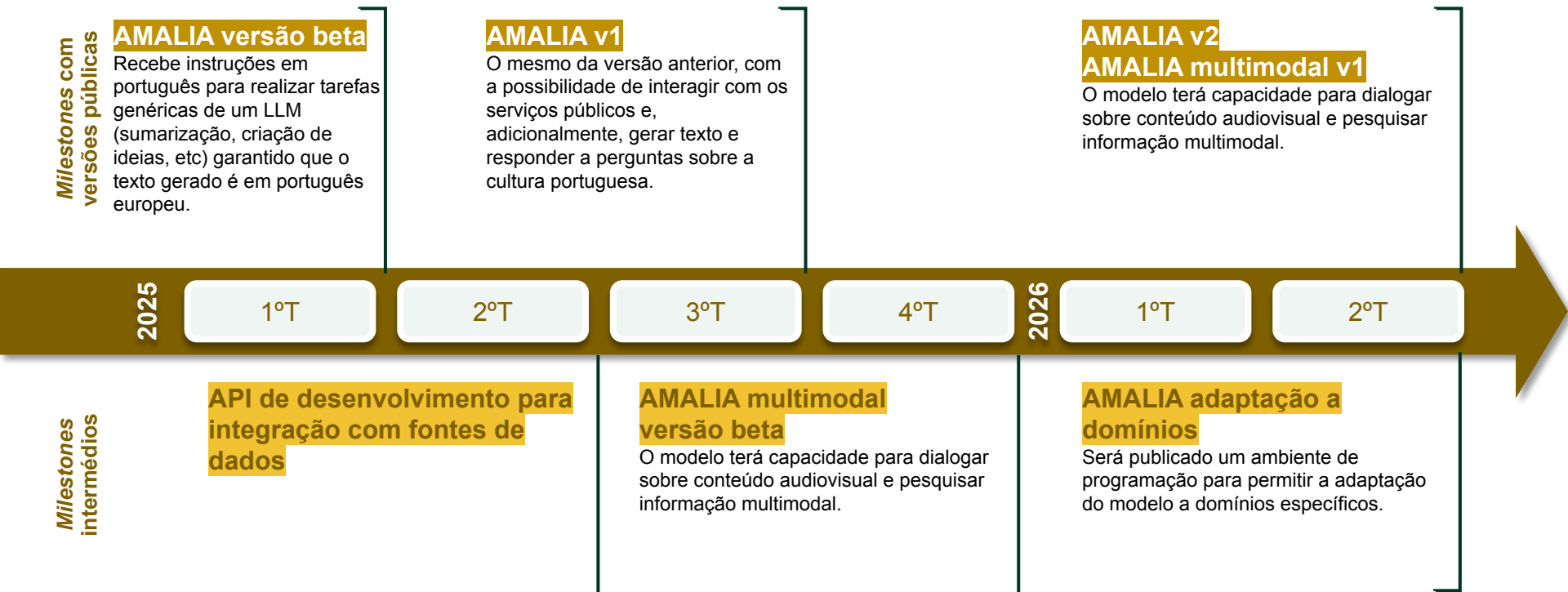
**UNIVERSIDADE D
COIMBRA**

U. PORTO

Março 2025: Primeira versão interna do AMALIA



Calendário de metas trimestrais





amália

LLM da Língua Portuguesa de Portugal

Obrigado.

<https://amalia.llm.pt/>

Processamento de dados

Existem dados para diferentes tarefas:

- Documentos com conhecimento enciclopédico.
- Documentos sobre a atualidade.
- Livros sobre cultura Portuguesa.
- Exames que exigem capacidade de raciocínio dedutivo ou conhecimento.
- Corpus paralelos para tradução.
- Dados com exemplos de sumarização.

A qualidade destes textos variam bastante e acabam por ser filtrados de acordo.



Treino e alinhamento do modelo

O treino é dividido em duas fases:

- O pré-treino consiste em treinar o modelo a prever palavras.
- O post-treino consiste em treinar o modelo a seguir instruções.

O treino por instruções irá dotar o modelo de mais capacidades mais sofisticadas. No limite as instruções podem ser uma sequência de ações.

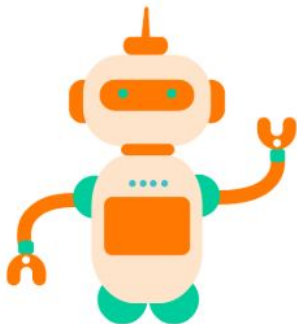
Ambas as equipas têm reconhecida experiência em ambos os tipos de treino.



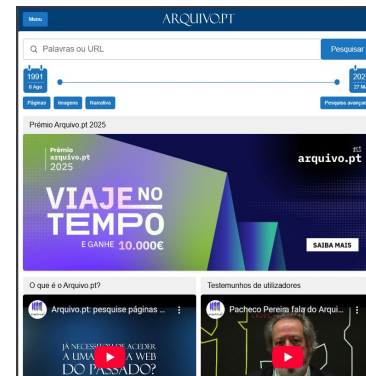
API Arquivo.pt

Acesso e processamento
automático da informação
preservada da Web

O que é uma API?



Website



API

```
1 // 20250327112712
2 // https://arquivo.pt/textsearch?
  q=fccn%20collection:AMP2&from=19970806&to=20140107&maxItems=16
3
4 {
5   "serviceName": "Arquivo.pt - the Portuguese web-archive",
6   "linkToService": "https://arquivo.pt",
7   "estimated_nr_results": 0,
8   "request_parameters": {
9     "offset": 0,
10    "dedupValue": 2,
11    "from": "19970806000000",
12    "to": "20140107000000",
13    "dedupField": "title",
14    "q": "fccn collection:AMP2",
15    "maxItems": 16
16  },
17  "response_items": [
18
19  ]
20 }
```

Os dados do Arquivo.pt

- Os dados do Arquivo.pt podem ser explorados de várias maneiras:
 - Pesquisa de texto e metadados
 - Pesquisa de imagens
 - Consulta de histórico de páginas arquivadas
- Um factor diferenciador do Arquivo.pt é possibilitar a exploração do **componente temporal** dos dados recolhidos
 - A mesma página é recolhida várias vezes em alturas diferentes

Pesquisa de texto - Pesquisa avançada

Termos de pesquisa: APDSI

Data entre 1 de Janeiro de 2003 e 1 de Janeiro de 2004

Formato: pdf

Menu

ARQUIVO.PT

Opções

Pesquisar

2003

1 Jan

2004

1 Jan

Páginas

Imagens

Narrativa

Pesquisa avançada

Cerca de 20 resultados desde 2003 até 2004

www.apdsi.pt/APDSI-Relatorio_2002.pdf

[PDF] **APDSI**

31 Julho 2003

Relatório de Actividades – 2002 Março 2003 RELATÓRIO DE ACTIVIDADES - 2002 **APDSI** ... DE ACTIVIDADES - 2002 **APDSI** – Associação para a Promoção e Desenvolvimento da Sociedade da ... Editorial "**APDSI** - Sociedade da Informação" • Prémio Personalidade do Ano no domínio da Sociedade da ...

www.apdsi.pt/APDSI-PA_2003_D5a.pdf

[PDF] **APDSI**

14 Setembro 2003

Plano de Actividades – 2003 Março 2003 PLANO DE DE ACTIVIDADES - 2003 **APDSI** ... estrangulamentos do Comércio e do Negócio Electrónicos em Portugal PLANO DE DE ACTIVIDADES - 2003 **APDSI** ... informático usado para fins de natureza social • Revista da **APDSI** • Internacionalizar a actividade da **APDSI** ...

Pesquisa de texto - API

<https://arquivo.pt/textsearch?q=APDSI&from=2003&to=20040101&type=pdf>

- **textsearch**
 - Serviço requisitado
- **q=APDSI**
 - Termos a pesquisar
- **from=2003**
 - resultados a partir do ano 2003
- **to=20040101**
 - resultados até 1 de Janeiro de 2004
- **type=pdf**
 - Ficheiros do formato pdf

```
{
  "serviceName": "Arquivo.pt - the Portuguese web-archive",
  "linkToService": "https://arquivo.pt",
  "estimated_nr_results": 20,
  "request_parameters": {
    "offset": 0,
    "dedupValue": 2,
    "from": "20030000000000",
    "to": "20040101000000",
    "type": [
      "pdf"
    ],
    "dedupField": "title",
    "q": "APDSI",
    "maxItems": 50
  },
  "response_items": [
    {
      "title": "APDSI",
      "originalURL": "http://www.apdsi.pt/APDSI-Relatorio_2002.pdf",
      "linkToArchive": "https://arquivo.pt/wayback/20030731233350/http://www.apdsi.pt/APDSI-Relatori",
      "tstamp": "20030731233350",
      "contentLength": 134217,
      "digest": "8e531f88573ce0d9b854bf8b65bf26f8",
      "mimeType": "application/pdf",
      "date": "1059694430",
      "linkToScreenshot": "https://arquivo.pt/screenshot?url=https%3A%2F%2Farquivo.pt%2FnoFrame%2Fre",
      "linkToNoFrame": "https://arquivo.pt/noFrame/replay/20030731233350/http://www.apdsi.pt/APDSI-R",
      "linkToExtractedText": "https://arquivo.pt/textextracted?m=http%3A%2F%2Fwww.apdsi.pt%2FAPDSI-R",
      "linkToMetadata": "https://arquivo.pt/textsearch?metadata=http%3A%2F%2Fwww.apdsi.pt%2FAPDSI-R"
```

Exemplo de utilização: conta-me histórias

conta-me histórias

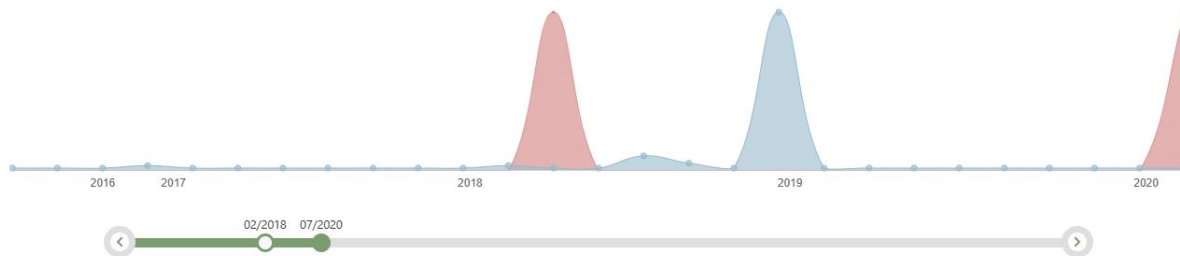
Crie uma narrativa de notícias sobre qualquer assunto

🔍 APDSI

[opções avançadas](#)

[Narrativa](#) [Termos relacionados](#)

Narrativa sobre **Apdsi** durante os últimos 10 anos



- Notícias relevantes entre **22/04/2018** e **17/07/2020**

- Geek Girls Portugal procuram raparigas entre os 13 e os 17 anos
- Aluno português ganha medalha de prata nas Olimpíadas de Informática no Japão
- Aluno português vence pela primeira vez medalha de prata nas Olimpíadas de Informática
- Como um piloto passa um chocolate para o banco de trás
- Boeing 757 aterra de lado em Bristol
- Older Ladies by Donnalou Stevens

contamehistorias.pt

Pesquisa de Imagens - Pesquisa avançada

Termo de pesquisa: APDSI

Data entre 1 de Janeiro de 2017 e a data de hoje

Imagens no formato JPG

Imagens provenientes do website apdsi.pt

Menu **ARQUIVO.PT** Opções

Q APDSI site:apdsi.pt type:jpg **Pesquisar**

2017 1 Jan 2025 1 Jan

Páginas Imagens Narrativa **Pesquisa avançada**

Cerca de 1.424 resultados desde 2017 até 2025

 → http://apdsi.pt/2020/a-ap... 12 Fevereiro 2020	 → http://apdsi.pt/2018/tertul... 10 Outubro 2018	 → http://apdsi.pt/2019/a-ap... 7 Fevereiro 2020	 → https://apdsi.pt/ 22 Abril 2021	 → http://apdsi.pt/2020/a-ap... 18 Janeiro 2020	 → http://apdsi.pt/produto/co... 24 Fevereiro 2020
 → https://apdsi.pt/2021/apd... 2 Junho 2021	 → http://apdsi.pt/2012/conf... 26 Julho 2018	 → http://apdsi.pt/2019/cam... 5 Dezembro 2019	 → http://apdsi.pt/2020/a-ap... 24 Fevereiro 2020	 → http://apdsi.pt/2019/apdsi... 4 Setembro 2019	 → https://apdsi.pt/2020/a-a... 24 Fevereiro 2020
 → http://apdsi.pt/produto/co... 5 Dezembro 2019	 → http://apdsi.pt/2019/a-ap... 4 Setembro 2019	 → http://apdsi.pt/2019/save... 22 Novembro 2019	 → http://apdsi.pt/2018/tertul... 10 Outubro 2018	 → http://apdsi.pt/produto/ap... 26 Fevereiro 2020	 → http://apdsi.pt/2018/asse... 16 Julho 2018

Pesquisa de Imagens - API

<https://arquivo.pt/imagesservice/?q=APDSI&from=2017&to=2025&type=jpg&siteSearch=apdsi.pt>

- **imagesservice**
 - Serviço requisitado
- **q=APDSI**
 - Termos a pesquisar
- **from=2017**
 - resultados a partir do ano 2017
- **to=2025**
 - resultados até 2025
- **type=jpg**
 - Imagens no formato JPEG
- **siteSearch=apdsi.pt**
 - Resultados do site apdsi.pt

```
{
  "serviceName": "Arquivo.pt - image search service.",
  "linkToService": "https://arquivo.pt/images.jsp",
  "linkToMoreFields": "https://arquivo.pt/imagesservice?q=APDSI&from=2017&to=2025&type=jpg&siteSearch=apdsi.pt&more=pageHost,matchingImages,safe",
  "nextPage": "https://arquivo.pt/imagesservice?q=APDSI&from=2017&to=2025&type=jpg&siteSearch=apdsi.pt&offset=50",
  "previousPage": "https://arquivo.pt/imagesservice?q=APDSI&from=2017&to=2025&type=jpg&siteSearch=apdsi.pt&offset=0",
  "totalItems": 1424,
  "numberOfResponseItems": 50,
  "offset": 0,
  "fieldReturnability": {
  },
  "responseItems": [
    {
      "imgDigest": "687e0582c32255ab670961a190425ce5ec4942806fd55884b2de692dc180754f",
      "imgSrc": "http://apdsi.pt/wp-content/uploads/2020/01/Conferencia-Conflitos-na-Era-da-IA-1024x480.jpg",
      "pageTitle": "A APDSI anuncia a Conferência “Os Conflitos na Era da Inteligência Artificial” - APDSI",
      "imgHeight": 480,
      "imgWidth": 1024,
      "imgMimeType": "image/jpeg",
      "pageURL": "http://apdsi.pt/2020/a-apdsi-anuncia-a-conferencia-os-conflitos-na-era-da-inteligencia-artificial/",
      "imgTstamp": "20200212064305",
      "pageTstamp": "20200210153925",
      "imgCaption": [
        "A APDSI anuncia a Conferência “Os Conflitos na Era da Inteligência Artificial” 16 Janeiro 2020 Notícia de APDSI, Áreas de Intervenção, Futuros, Futuros da Sociedade da Informação, Grupos",
        "A APDSI anuncia a Conferência “Os Conflitos na Era da Inteligência Artificial” Inscreva-se aqui A APDSI organiza a\entre ... Continue a ler A APDSI anuncia a Conferência “Os Conflitos na Era da Inteligência Artificial” APDSI 0 Comentários",
        "Conflitos na Era da IA APDSI 0 Comentários"
      ]
    }
  ]
}
```

Exemplo de utilização: Imaginarium



imagination.pages.dev/

Consulta de histórico de páginas arquivadas

Menu

ARQUIVO.PT

Q apdsi.pt

Pesquisar

1991 6 Ago

2025 27 Mar

Páginas

Imagens

Narrativa

Pesquisa avançada

Cerca de 1.113 resultados desde 1991 até 2025

Tabela

Lista

2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
2 Ago	24 Jan	31 Jan	5 Fev	6 Jan	1 Jan	21 Out	25 Mai	2 Jun	19 Mai	21 Jan	5 Nov	23 Set	11 Abr	5 Fev	14 Fev
25 Nov	20 Mar	1 Abr	8 Fev	8 Jan	5 Jan		25 Set			31 Jan	8 Nov	29 Set	23 Abr	10 Fev	17 Fev
28 Nov	28 Mar	5 Jun	12 Fev	11 Jan	10 Jan		21 Dez						11 Set	30 Mai	21 Fev
	25 Abr	10 Jun	6 Mar	12 Jan	15 Jan								19 Set	6 Jun	7 Jun
	7 Jun	12 Jun	8 Mar	15 Jan	26 Jan								16 Nov	31 Out	31 Jul
	24 Jul	27 Ago	4 Abr	30 Jan	2 Fev								23 Nov	6 Nov	6 Ago
	21 Set	2 Set	11 Abr	2 Fev	21 Mar										7 Dez
	10 Out	24 Set	29 Mai	6 Fev	4 Abr										12 Dez
	23 Nov	6 Dez	18 Jun	7 Fev	7 Jun										
	26 Nov		20 Jun	8 Fev	10 Jun										
			24 Jun	15 Fev	12 Set										
			22 Jul	16 Fev											

Menu

ARQUIVO.PT

Q apdsi.pt

Pesquisar

1991 6 Ago

2025 27 Mar

Páginas

Imagens

Narrativa

Pesquisa avançada

Cerca de 1.113 resultados desde 1991 até 2025

Tabela

Lista

2002	3 versões
2003	10 versões
2004	9 versões
2005	22 versões
2006	31 versões
2007	11 versões
Janeiro	5 versões

1 Janeiro 04h35, 2007

5 Janeiro 20h53, 2007

10 Janeiro 15h53, 2007

15 Janeiro 17h20, 2007

Consulta de histórico de páginas arquivadas

Menu

ARQUIVO.PT

Q apdsi.pt

Pesquisar

1991

6 Ago

2025

27 Mar

Páginas

Imagens

Narrativa

Pesquisa avançada

Cerca de 1.113 resultados desde 1991 até 2025

Tabela

Lista

2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
2 Ago	24 Jan	31 Jan	5 Fev	6 Jan	1 Jan	21 Out	25 Mai	2 Jun	19 Mai	21 Jan	5 Nov	23 Set	11 Abr	5 Fev	14 Fev
25 Nov	20 Mar	1 Abr	8 Fev	8 Jan	5 Jan		25 Set			31 Jan	8 Nov	29 Set	23 Abr	10 Fev	17 Fev
28 Nov	28 Mar	5 Jun	12 Fev	11 Jan	10 Jan		21 Dez						11 Set	30 Mai	21 Fev
	25 Abr	10 Jun	6 Mar	12 Jan	15 Jan								19 Set	6 Jun	7 Jun
	7 Jun	12 Jun	8 Mar	15 Jan	26 Jan								16 Nov	31 Out	31 Jul
	24 Jul	27 Ago	4 Abr	30 Jan	2 Fev								23 Nov	6 Nov	6 Ago
	21 Set	2 Set	11 Abr	2 Fev	21 Mar										7 Dez
	10 Out	24 Set	29 Mai	6 Fev	4 Abr										12 Dez
	23 Nov	6 Dez	18 Jun	7 Fev	7 Jun										
	26 Nov		20 Jun	8 Fev	10 Jun										
			24 Jun	15 Fev	12 Set										
			22 Jul	16 Fev											

pt uk

ARQUIVO.PT

Início

Sítios de interesse

Mapa

Ajuda

Directo

Contactos

Delimitação de Responsabilidade

Inscreva-se na nossa Associação

Individual

Collectivo

Inscreva-se nos Grupos de Estudo da APDSI

APDSI

Sócios

Actividades

Prémios e Homenagens

Parcerias

Publicações



Atalhos de acesso ao menu: ALT+ Letra sublinhada

Carta Aberta "Fazer crescer o mercado da SI em Portugal"

A APDSI apresentou hoje uma nova Carta Aberta. O objectivo da iniciativa foi contribuir para o crescimento e desenvolvimento da Sociedade da Informação na sociedade portuguesa, analisando os seus factores de constrangimento e as suas oportunidades. [30-05-2007]

ver mais

Olimpíadas da Informática

Chegou ao fim a 19.ª edição das Olimpíadas Nacionais de Informática (ONI'2007), mais uma vez promovidas pela APDSI. [23-05-2007]

ver mais

Faleceu António João Simões Monteiro

Faleceu no passado dia 18 de Maio António João Simões Monteiro, Director-Geral e sócio fundador. A APDSI presta uma sentida homenagem ao Eng.º António João Simões Monteiro e acompanha os seus familiares neste momento de luto. [23-05-2007]

Apresentação Pública do Estudo "Identidade Digital"

A adopção da identidade digital, como condição essencial para uma adopção plena do mundo digital, tem subjacente um número alargado de perigos que podem ser contornados se for dada resposta a um conjunto de princípios fundamentais, conclui o estudo Identidade Digital elaborado por um grupo de trabalho da APDSI - Associação para a Promoção e Desenvolvimento da Sociedade da Informação e ontem apresentado publicamente em Lisboa. [19-04-2007]

ver mais

Conferência "e-Inclusão - Um desafio para Portugal"

No âmbito das comemorações dos 10 anos da publicação do Livro Verde para a Sociedade da Informação e após cinco anos de actividade associativa, a APDSI realizou hoje, dia 11 de Abril, na

[arquivo.pt/textsearch?versionHistory=apdsi.pt](#)

Outra forma de consultar o histórico - CDX API

```
pt,apdsi)/ 20020802073300 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "ZVM3GR6CHCT5GCRD40GEIRVMQ4LKHRTF", "length":
pt,apdsi)/ 20021125214442 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "6303LJFRTATXY5NF6KCJ3HAU05JSSFGV", "length":
pt,apdsi)/ 20021128195927 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "6303LJFRTATXY5NF6KCJ3HAU05JSSFGV", "length":
pt,apdsi)/ 20030124211351 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "3H3PPMDF67IR2VXWQ2XRNEATX50ONTBQ", "length":
pt,apdsi)/ 20030320021622 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "QCJ3JXFK25IVBPCX4ZL3EILB2H05PWA0", "length":
pt,apdsi)/ 20030328172315 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "B4GZ0YRUPZTNGSMLXQMP4Y5E2YOKGBX7", "length":
pt,apdsi)/ 20030425212318 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "SRLXUVITXG3TWY757WKUEROX620RMTOT", "length":
pt,apdsi)/ 20030607135157 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "CMEV5IXASQACPCB3UFQ0SIAYDSICBY02", "length":
pt,apdsi)/ 20030724082503 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "GTQ7MDXDUACCYA0F7TEBRMEYCEXDJGKN", "length":
pt,apdsi)/ 20030921222349 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "GTQ7MDXDUACCYA0F7TEBRMEYCEXDJGKN", "length":
pt,apdsi)/ 20031010061938 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "26X62WVEYA50EFKJP74ZSXFHH50UD6M", "length":
pt,apdsi)/ 20031123213605 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "DPWQA4AWV2ISN40073BYP7C5F2CDEQ20", "length":
pt,apdsi)/ 20031126000717 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "DPWQA4AWV2ISN40073BYP7C5F2CDEQ20", "length":
pt,apdsi)/ 20031126195130 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "DPWQA4AWV2ISN40073BYP7C5F2CDEQ20", "length":
pt,apdsi)/ 20040131064511 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "UOEFZKY4R2VYU5PEKZAUOTFU0SZZIE27", "length":
pt,apdsi)/ 20040401123404 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "VF7I2MB4POWZUFDYWG3SLB2XV07Q72W", "length":
pt,apdsi)/ 20040605091744 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "3B7VWMHSX2CXVM6PFYQ5GFGU2CTWGSQ4S", "length":
pt,apdsi)/ 20040610235650 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "S7MMKWBXYT4ZCKXFJ5I0STHKHJHHADBU", "length":
pt,apdsi)/ 20040612122853 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "S7MMKWBXYT4ZCKXFJ5I0STHKHJHHADBU", "length":
pt,apdsi)/ 20040827204227 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "TF6GBHARFYGJNYIHD8EFPHGHCJWBLI7B", "length":
pt,apdsi)/ 20040902023137 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "TF6GBHARFYGJNYIHD8EFPHGHCJWBLI7B", "length":
pt,apdsi)/ 20040924030314 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "UM6JR6E6WGFVRV6BKUDE7B07ESRDWRBIW", "length":
pt,apdsi)/ 20041206224055 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "GI0DW6CJ30XNMWRXQK8G42U56NC4CAMU2", "length":
pt,apdsi)/ 20050205064254 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "32HNLYP60FOSLFY3M2FK7XWMPKC64S5C", "length":
pt,apdsi)/ 20050208213141 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "32HNLYP60FOSLFY3M2FK7XWMPKC64S5C", "length":
pt,apdsi)/ 20050212224516 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "RTVITQTVDQZ354PXCNSTX4KBDGBRMQ6J", "length":
pt,apdsi)/ 20050306090619 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "BDIWA4GHXVOMLWZ07PHNRTC4ZG55FB7V", "length":
pt,apdsi)/ 20050308131546 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "BDIWA4GHXVOMLWZ07PHNRTC4ZG55FB7V", "length":
pt,apdsi)/ 20050404010050 {"url": "http://www.apdsi.pt:80/", "mime": "text/html", "status": "200", "digest": "JTLGJAH24YBVCB4R6MQIKKBVC66BPTC", "length":
pt,apdsi)/ 20050411190511 {"url": "http://www.apdsi.pt/", "mime": "text/html", "status": "200", "digest": "K3BPTZUQGAKE2E66F20HVVERR7CZWQZT", "length":
```

arquivo.pt/wayback/cdx?url=apdsi.pt

Consulta em massa – download dos CDX

Index of /cdxj

 [ICO]	Name	Last modified	Size	Description
	[PARENTDIR] Parent Directory		-	
	[] AWP1.cdxj	2023-02-03 13:01	18G	
	[] AWP2.cdxj	2023-02-03 13:04	20G	
	[] AWP3.cdxj	2023-02-03 13:07	21G	
	[] AWP4.cdxj	2023-02-03 13:25	30G	
	[] AWP5.cdxj	2023-02-03 14:08	49G	
	[] AWP6.cdxj	2023-02-03 14:44	52G	
	[] AWP7.cdxj	2023-02-03 14:23	39G	
	[] AWP8.cdxj	2022-12-04 02:12	36G	
	[] AWP9.cdxj	2023-02-03 14:27	40G	
	[] AWP10.cdxj	2023-02-03 13:36	34G	
	[] AWP11.cdxj	2023-02-03 13:29	32G	
	[] AWP12.cdxj	2023-02-03 14:00	45G	
	[] AWP15.cdxj	2023-02-03 14:38	65G	
	[] AWP16.cdxj	2023-02-03 15:59	104G	

arquivo.pt/cdxj

Colecções do Arquivo.pt

List of the collections preserved by Arquivo.pt (publicly available)

Collections About							
Collection ID	Description	Collection Start date	Collection End date	Total files (Number of URLs collected)	Total seeds (Preserved sites)	Total volume of WARCS files (TB)	Collection in Production? (URL, Page and Image indexed)
AWP1	1st complete crawl of the Portuguese web, mainly from the .PT domain, in 2008.	2008-02-12	2008-03-06	56,046,288	154,787	1.60	TRUE
AWP2	2nd complete crawl of the Portuguese web, mainly from the .PT domain, in 2008.	2008-03-11	2008-05-30	48,718,404	-	1.60	TRUE
AWP3	3rd complete crawl of the Portuguese web, mainly from the .PT domain, in 2008.	2008-10-21	2008-12-10	51,863,006	193,294	2.00	TRUE
AWP4	4th complete crawl of the Portuguese web, mainly from the .PT domain, in 2009.	2009-05-01	2009-05-31	68,776,707	366,880	2.50	TRUE
AWP5	5th complete crawl of the Portuguese web, mainly from the .PT domain, in 2009.	2009-10-01	2009-10-31	119,135,566	373,323	3.80	TRUE
AWP6	6th complete crawl of the Portuguese web, mainly from the .PT domain, in 2009.	2009-12-01	2009-12-31	118,810,364	340,018	3.50	TRUE
AWP7	7th complete crawl of the Portuguese web, mainly from the .PT domain, in 2010.	2010-05-01	2010-05-31	87,988,812	389,957	2.90	TRUE
AWP8	Incremental crawl of the Portuguese web, mainly from the .PT domain, in 2010. The AWP8 crawl is incremental because it was performed using DeDuplicator (http://landsbokasafn.github.io/DeDuplicator/) taking the content of AWP7 as baseline. Thus, the files that remained unchanged from the AWP7 complete crawl were not archived (duplicated) on the AWP8 incremental crawl.	2010-08-01	2010-08-31	75,771,317	411,562	1.90	TRUE
AWP9	Incremental crawl of the Portuguese web, mainly from the .PT domain, in 2011. The AWP9 crawl is incremental because it was performed using DeDuplicator (http://landsbokasafn.github.io/DeDuplicator/) taking the content of AWP7 as baseline. Thus, the files that remained unchanged from the AWP7 complete crawl were not archived (duplicated) on the AWP9 incremental	2011-01-20	2011-03-22	81,114,575	473,588	2.10	TRUE

arquivo.pt/collections

Um exemplo de utilização: Modelo LLM

Glória, um modelo para a língua portuguesa

Um dos projetos que utilizou o Arquivo.pt para obter grande quantidade de texto denomina-se Glória e é um modelo de linguagem em larga escala ([LLM, Large Language Model](#)) focado na língua portuguesa europeia.

“Apesar da abundância de LLMs para muitas línguas de recursos elevados, a disponibilidade de tais modelos continua a ser limitada para o português europeu”, como explicam os autores do Glória, Ricardo Lopes, João Magalhães, David Semedo, investigadores da [Faculdade de Ciência e Tecnologia da Universidade Nova de Lisboa](#), no seu artigo [Glória – A Generative and Open Large Language Model for Portuguese](#).

Questões?

arquivo.pt/api

contacto@arquivo.pt