

Prémio Arquivo.pt

Descrição Sumária do Trabalho

Identificação

- **Título:** Arquivo do Parlamento: Artigos de legislaturas passadas disponíveis para o futuro.
- **Área temática:** Ciências Sociais; Ciência Política; Tecnologias da Informação.
- **Candidato:** Tiago Santos
- **Email:** mail@tiago-santos.eu
- **Website:** <https://arquivo-parlamento.pt>

Descrição do Trabalho

Num universo mediático cada vez mais fragmentado onde a partilha diária de milhões de notícias nas redes sociais se tornou um lugar-comum e as *fake news* desafiam o que tomamos por verdadeiro, torna-se cada vez mais necessário criar espaços que resistam a essa voragem de fragmentação e nos permitam navegar no oceano de informação. Nesse sentido, o Arquivo do Parlamento surge como um repositório aberto que procura compilar e dar visibilidade às pequenas e grandes narrativas sobre a vida política portuguesa.

Recorrendo às páginas *web* arquivadas pelo Arquivo.pt e aos dados abertos do Parlamento.pt, o Arquivo do Parlamento agrega notícias e artigos de opinião sobre a atividade do Parlamento, as suas legislaturas, grupos parlamentares e deputados. De modo a permitir análises mais completas e heterogéneas, considerámos notícias e artigos de opinião produzidos pelos *media* tradicionais, como por exemplo o jornal *Público*, mas também aqueles que são criados pelos cidadãos, nomeadamente os conteúdos disponibilizados em blogues pessoais.

A componente técnica compreende quatro fases distintas:

- 1 Recolha dos dados sobre todas as legislaturas desde 1996 (VII - XIV) a partir dos dados abertos disponíveis em Parlamento.pt;
- 2 Extração das páginas *web* do Arquivo.pt através da pesquisa do nome dos grupos parlamentares e deputados;

- 3 Extração do título, texto e palavras-chave dos artigos, identificação dos grupos parlamentares e deputados e inserção desta informação na base de dados;
- 4 Desenvolvimento do *website*, com a ajuda da *framework open-source Django*, de modo a tornar a informação acessível a toda a comunidade.

The screenshot displays the 'Arquivo do Parlamento' website. At the top, a navigation menu includes 'Legislaturas', 'Grupos parlamentares', 'Deputados', 'Media', 'Tópicos', and 'Artigos'. The main heading is 'Arquivo do Parlamento.' followed by the text 'Artigos de legislaturas passadas disponíveis para o futuro.' Below this, a summary table shows the following counts: LEGISLATURAS: 8, GRUPOS PARLAMENTARES: 10, DEPUTADOS: 1363, MEDIA: 7625, TÓPICOS: 12140, and ARTIGOS: 217693.

The main content area is titled 'XIV Legislatura 25-10-2019 | 28-03-2022'. It features a grid of legislative groups with their respective logos, names, and statistics:

Logo	Party Name	Deputados	Artigos
	PS Partido Socialista	129	9161
	PSD Partido Social Democrata	85	6303
	BE Bloco De Esquerda	22	3326
	PCP Partido Comunista Português	12	3533
	CDS-PP Centro Democrático Social - Partido Popular	8	1483
	PAN Pessoas-Animais-Natureza	5	1942
	PEV Partido Ecologista "Os Verdes"	2	1078
	CH Chega	2	1529
	IL Iniciativa Liberal	1	1005
	L Livre	1	640

Below the grid, a list of previous legislatures is shown with expandable arrows:

- XIII Legislatura 23-10-2015 | 24-10-2019
- XII Legislatura 20-06-2011 | 22-10-2015
- XI Legislatura 15-10-2009 | 19-06-2011
- X Legislatura 10-03-2005 | 14-10-2009
- IX Legislatura 05-04-2002 | 09-03-2005
- VIII Legislatura 25-10-1999 | 04-04-2002
- VII Legislatura 27-10-1995 | 24-10-1999

Figura 1: Página inicial Arquivo do Parlamento (<https://arquivo-parlamento.pt>)

Objetivos

O principal objetivo Arquivo do Parlamento passa por colocar ao dispor de todos uma ferramenta de investigação poderosa e facilmente utilizável. Este *website* compila e disponibiliza dados que, normalmente, são inacessíveis ao utilizador comum, devido à sua elevada dispersão e à complexidade técnica dos processos de tratamento e análise de *Big Data*. Possibilita, por exemplo, visualizar os partidos e deputados mais visados pelos media em cada legislatura, os *media* mais presentes ou os tópicos mais salientes.

Este projeto foi também pensado de forma a poder facilmente servir de material de estudo e pesquisa em diferentes áreas de investigação, como as Ciências Sociais, a Ciência Política ou as Tecnologias da Informação. Através do botão “descarregar ficheiro CSV” (Comma Separated Values) é possível aos investigadores ter acesso à informação num formato comum de ficheiro.

Resultados Atingidos

O Arquivo do Parlamento reúne informação de 8 Legislaturas, 10 Grupos Parlamentares, 1363 deputados produzida por 7625 *Media*, categorizada em 12140 Tópicos, num total de 217693 Artigos recolhidos e processados.

A informação está organizada de forma hierárquica em Legislaturas, Grupos Parlamentares, Deputados, *Media*, Tópicos e Artigos. Desta forma, é possível ter diversos tipos e níveis de análise em simultâneo. Por exemplo, ao navegarmos por Legislatura estamos a filtrar os artigos temporalmente, o que permite não só visualizar os grupos parlamentares, deputados, *media*, tópicos mais relevantes, mas também navegar dentro destas categorias e descarregar os artigos a elas pertencentes. Por outro lado, ao acedermos aos itens de menu Grupos Parlamentares, Deputados, *Media*, Tópicos e Artigos, temos acesso a toda informação relativa a estas categorias independentemente do seu contexto temporal, o que permite análises mais alargadas. Dentro de cada uma das páginas de listagem de artigos é ainda possível filtrar de diversas formas, consoante o contexto, de modo a tornar a informação mais acessível.

XIV Legislatura

25-10-2019 | 28-03-2022

PARTIDOS DEPUTADOS MEDIA TÓPICOS ARTIGOS
10 267 1766 9325 30487

Grupos parlamentares

 PS Partido Socialista DEPUTADOS: 129 ARTIGOS: 9161	 PSD Partido Social Democrata DEPUTADOS: 85 ARTIGOS: 6303	 BE Bloco De Esquerda DEPUTADOS: 22 ARTIGOS: 3326	 PCP Partido Comunista Português DEPUTADOS: 12 ARTIGOS: 3533
 CDS-PP Centro Democrático Social - Partido Popular DEPUTADOS: 8 ARTIGOS: 1483	 PAN Pessoas-Animais-Natureza DEPUTADOS: 5 ARTIGOS: 1942	 PEV Partido Ecologista "Os Verdes" DEPUTADOS: 2 ARTIGOS: 1078	 CH Chega DEPUTADOS: 2 ARTIGOS: 1529
 IL Iniciativa Liberal DEPUTADOS: 1 ARTIGOS: 1005	 L Livre DEPUTADOS: 1 ARTIGOS: 640		

Deputados populares

 António Costa PS ARTIGOS: 17080	 Rui Rio PSD ARTIGOS: 11690	 André Ventura CH ARTIGOS: 7571	 Joaquim Katar Moreira L ARTIGOS: 5945
 Mário Centeno PS ARTIGOS: 4790	 Catarina Martins BE ARTIGOS: 4616	 André Silva PAN ARTIGOS: 4495	 Jerónimo De Sousa PCP ARTIGOS: 4072
 Pedro Nuno Santos PS ARTIGOS: 3207	 João Cotrim De Figueiredo L ARTIGOS: 2866	 José Luís Ferreira PEV ARTIGOS: 2795	 Cristina Rodrigues PAN ARTIGOS: 2758
 Mariana Mortágua BE ARTIGOS: 2737	 João Oliveira PCP ARTIGOS: 2635	 Telmo Correia CDS-PP ARTIGOS: 2592	 Pedro Filipe Soares BE ARTIGOS: 2492

Media populares

Sapo ARTIGOS: 6871	Expresso ARTIGOS: 2653	Observador ARTIGOS: 2230	Sicnoticias ARTIGOS: 665
Juramentosembandeira ARTIGOS: 540	Corporacoes ARTIGOS: 540	Portugalospequeninos ARTIGOS: 408	Alguresaqui ARTIGOS: 363
Comqueentao ARTIGOS: 357	Tvi24 ARTIGOS: 351	Impossiveldeaturar ARTIGOS: 331	Pedrogaosaopedro ARTIGOS: 317
Gotadeagua53	Opafuncao	Fragmentosdebondade	Publico

Figura 2 Página da XIV Legislatura

Originalidade e caráter inovador

O Arquivo do Parlamento é o primeiro projeto a reunir os dados do Parlamento.pt com as páginas *web* arquivadas no Arquivo.pt de modo a agregar notícias e artigos de opinião sobre a atividade do Parlamento, as suas legislaturas, grupos parlamentares e deputados. Deste modo, é disponibilizada informação sobre a vida política portuguesa que, de outro modo, seria inacessível aos investigadores e utilizadores comuns.

Impacto social

Para lá da utilidade científica, o Arquivo do Parlamento foi idealizado e desenhado de forma a ser facilmente utilizado por utilizadores comuns (não técnico-científicos), sem uma formação específica. Como já anteriormente mencionado, para além de facilitar o acesso à informação, este *website* serve como ferramenta de combate à desinformação e *fake news*, ajudando os cidadãos a gerir o imenso volume de informação produzido diariamente.

Funciona também como um motor de pesquisa que procura identificar e mapear os discursos produzidos acerca da vida política portuguesa e seus protagonistas, ajudando a responder a questões que os cidadãos tenham sobre os mesmos. Ao fazê-lo, atribui a mesma importância ao discurso produzido pelos *media* tradicionais, e ao que é produzido pelos cidadãos comuns. Salientamos, por isso, o contributo deste *website* para dar visibilidade a discursos frequentemente excluídos da construção das grandes narrativas sobre estas temáticas.

Por fim, importa mencionar o papel deste *website* na valorização e preservação do trabalho jornalístico e dos conteúdos escritos sobre a vida política portuguesa, que representa as perspetivas que o país tem sobre a mesma ao longo do tempo.

Impacto científico

Este projeto facilita, desde logo, o acesso à informação – e especificamente a *Big Data* – e permite a abertura de novas linhas de investigação nos domínios das Ciências Sociais, Ciência Política ou Tecnologias da Informação, entre outros. Futuramente serão disponibilizadas APIs para consulta pública dos dados e o código será disponibilizado em acesso aberto.

Relevância da utilização do Arquivo.pt

O Arquivo.pt é o eixo estruturante deste projeto. Sem as API's do Arquivo.pt, não conseguiríamos desenvolver de forma tão completa e consistente este trabalho, dado que a maior parte dos artigos que disponibiliza já não está *online*, principalmente os mais antigos. Não fosse o Arquivo.pt, esta informação, que é parte integrante da nossa memória coletiva, estaria irremediavelmente perdida. Neste sentido, louvamos o esforço feito até aqui para manter este arquivo da informação publicada na *web* portuguesa esteja acessível às gerações futuras.

Recursos complementares

- Dados abertos em Parlamento.pt, <https://www.parlamento.pt/Cidadania/Paginas/DadosAbertos.aspx>: Informação base relativa à Assembleia da República, que integra dados sobre Legislaturas, Sessões Legislativas, Grupos Parlamentares, Círculos Eleitorais e Deputados.