

Prémio Arquivo.pt

Descrição Sumária do Trabalho

Identificação

- Título: Existo – Representatividade das mulheres artistas na imprensa nacional
- Área temática: Processamento de Língua Natural, Cultura, Artes, Informática
- Candidato: Cláudia Sofia Sevivas Ribeiro, Miguel Boavida
- Email: claudia.sevivas@universidadeeuropeia.pt,
miguel.boavida@universidadeeuropeia.pt
- Website: <https://www.existo.pt>

Descrição do Trabalho

Num mundo onde as tecnologias já são o que fazemos, o que escrevemos, o que pensamos... será que também são onde existimos? onde somos? onde estamos representados? Este projeto começou com essa premissa. Se é onde estamos representados, para o mundo, para a memória, para o conhecimento acumulado que nos torna humanos, onde estão as artistas mulheres portuguesas? Elas estão, são, existem?

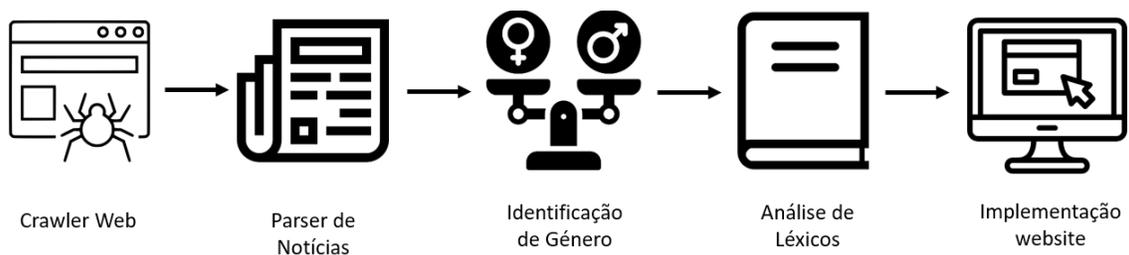
Pensámos, será que a inteligência artificial, tão diligente, que nos filtra o mundo e nos guia na forma de pensar, consegue ser imparcial, inclusivo, revelar o que existe para além das sombras, dos filtros, do que escolhemos ver?

Quisemos então tentar medir essa ausência, de uma forma automática que nos permita explorar de que forma essa ausência se sente nas várias artes e ao longo do tempo. Por isso o projeto se chama Existo, porque existir é viver, estar vivo, ... é a não ausência.

Este projeto centra-se na criação de um *website*, acreditamos que seja o primeiro a tentativamente disponibilizar uma análise da representação e visibilidade das artistas portuguesas. No contexto desta prova de conceito usamos o prestigiado suplemento Ípsilon do jornal Público. Paralelamente a essa análise é também disponibilizada informação sobre as artistas mais mencionadas, as notícias onde estas foram mencionadas ou citadas, a distribuição por área geográfica e a agregação da análise de léxico (i.e., palavras mais repetidas) do texto da notícia. O trabalho foi desenvolvido por dois docentes do IADE – Creative University.



O desenvolvimento e implementação do projeto pode ser descrito em cinco fases:



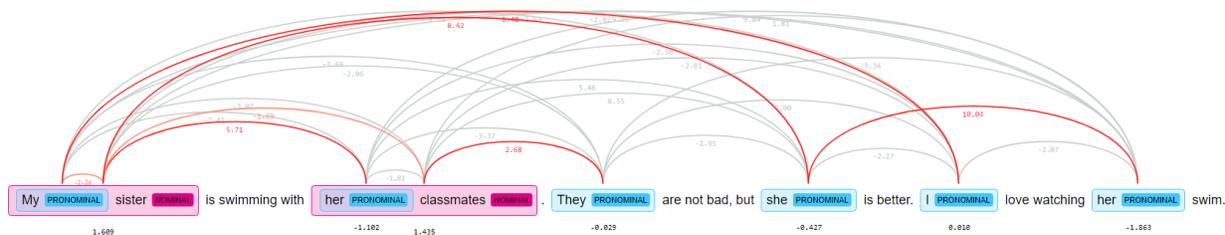
Crawler Web: Criação de scripts para executar *data mining* de notícias do suplemento Ípsilon do jornal Público entre 2014 e 2021.

Parser de Notícias: Programação de ferramentas para fazer o *parsing* e limpeza de informação relevante dos dados obtidos de mais de 12 mil notícias em formato HTML e JSON. É também neste *parser* que a informação é estruturada e persistida numa base de dados MongoDB.

Identificação de Género: Esta é a fase central do projeto, onde se identifica quantas referências masculinas, femininas e ambas existem no corpo da notícia. Para isso, usa-se a técnica de reconhecimento de entidades nomeadas. A partir desta técnica é possível extrair de cada frase/parágrafo os seguintes identificadores: (PER – pessoa), ORG (organização), LOC (localização) entre outras, como demonstrado no seguinte exemplo:

Jeff Bezos **PERSON** founded e-commerce giant Amazon **ORG** in 1994 **DATE** out of his garage in Seattle **LOC**. He will step down as CEO to become executive chairman in July 2021 **DATE**.

Nas situações em que não existem pessoas mencionadas usou-se a técnica de resolução de co-referências (exemplificado na imagem em baixo), de forma a desambiguar pessoas mencionadas e também o género.



Quando existem apenas pessoas mencionadas e não existem co-referências, o género é extraído a partir da análise do nome. Para isso implementou-se um conjunto de funções baseadas na API Namsorⁱ e Wiki-Gendersortⁱⁱ.

Análise de Léxico: Para cada rubrica, é feita uma análise integrada das palavras mais repetidas no corpo das notícias. Para tal usou-se o SentiLex-PTⁱⁱⁱ, que é uma lista de lemas composto por 82 347 formas flexionadas, associadas aos respetivos lemas e estão repartidas em: 16 863 adjetivos, 1 280 substantivos, 29 504 verbos. Cada lema tem também associado a polaridade: 1 (positivo), 0 (negativo) e -1 (negativo).

Implementação Website: Para a implementação do *website* usou-se tecnologia *MEAN Stack* que é o conjunto de quatro tecnologias *open-source* de desenvolvimento web, em que *MEAN* representa as tecnologias de MongoDB, ExpressJS, Angular e NodeJS.

Objetivos

Os principais objetivos do projeto são:

- Mostrar a representatividade e visibilidade dada às mulheres artistas na imprensa nacional, a distribuição das notícias geograficamente, e as palavras mais repetidas, em diferentes géneros artísticos, recorrendo a *infographics*;
- A partir de técnicas de processamento de língua natural, extrair todas as artistas mencionadas desde 2014 de forma a construir uma base de dados que poderá enriquecer outros projetos relacionados e que irá ser disponibilizada para uso e consulta pública;
- Fornecer uma ferramenta de consulta onde potenciais utilizadores e investigadores podem aceder a todas as notícias de uma determinada artista, género artístico ou ano. Paralelamente analisar tendências associadas à visibilidade das mulheres artistas.

Este trabalho irá ser expandido de forma a incluir género não binário e artistas da diáspora. Para além disso, pretende-se que o projeto seja alargado e atualizado anualmente, conforme novo material seja disponibilizado pelo Arquivo.pt e novas fontes, como por exemplo, o site CoffeePaste – a plataforma da comunidade das artes (<https://coffeepaste.com/>).

Resultados Atingidos

Este projeto conta com 12.000 mil artigos processados, os quais constituem os artigos presentes no arquivo.pt entre 2014 e 2021. A partir desses artigos foi possível identificar automaticamente a partir das ferramentas desenvolvidas e técnicas de NLP utilizadas cerca de seguintes 300 artistas portuguesas.

Arquitetura	Inês Lobo, Graça Fonseca, Patrícia Barbas, Teresa Novais, Ana Tostões, Dalila Rodrigues, Catarina Mourão, Ana Vaz Milheiro, Sofia Aleixo, Magda Seifert, Guta Moura Guedes, Maria Cardoso, Paula Torgal, Paula Silva, Leonor Antunes, Leonor Picão, Célia Gomes
Cinema	Teresa Villaverde, Cíntia Gil, Pandora da Cunha Telles, Salomé Lamas, Anabela Moreira, Rita Azevedo Gomes, Mónica Santos, Catarina Mourão, Ana Moreira, Beatriz Batarda, Catarina Alves Costa, Margarida Gil, Ana Isabel Strindberg, Joana de Verona, Joana Pimenta, Susana Nobre, Maria de Medeiros, Maria João Seixas, Maria do Céu Guerra, Laura Gonçalves, Teresa Madruga, Maria João Luís, Leonor Noivo, Renata Pinheiro, Catarina Wallenstein, Rossana Torres, Laura Marques, Isabel Ruth, Patrícia Figueiredo, Alice Guimarães, Victória Guerra, Mónica Baptista, Bárbara Virgínia, Maria Cabral, Maria Pinto, Maria João Bastos, Ana Maria Pinto, Madalena Iglésias, Inês de Medeiros, Filipa César, Júlia Palh, Joana Bárcia, Rita Lino, Marta Mateus, Victória Guerra, Júlia Buisel, Leonor Silveira, Clara Riedenstein, Rita Blanco, Cristèle Alves Meira, Lúcia Prancha, Mariana Gaivão, Anabela Moutinho, Tânia Dinis, Beatriz Costa, Mariana Caló, Inês Oliveira, Margarida Lucas, Crista Alfaiate, Luísa Cruz, Sofia Bost, Ana Brito e Cunha, Carla Bolito, Marta Madureira, Leonor Teles, Filipa Reis, Márcia Breia, Catarina Ruivo, Alexandra Ramires, Laura Paredes, Margarida Moreira, Carla Maciel, Joana Barrios, Maria Leite, Lia Gama, Catarina Vasconcelos
Dança	Vera Mantero, Olga Roriz, Marlene Monteiro Freitas, Clara Andermatt, São Castro, Lia Rodrigues, Joana Castro, Joana Providência, Sofia Campos, Elizabeth Francisca, Teresa Alves da Silva, Paula Garcia, Marcela Levi, Cristina Planas Leitão, Mariana Tengner Barros, Madalena Victorino, Catarina Félix, Daniela Cruz, Maurícia Neves, Inês Barahona, Inês Jacques, Paula Sá Nogueira, Carlota Lagido, Inês Campos, Vânia Rovisco, Sara Carinhas, Dina Magalhães, Filipa Peraltinha, Ana Rocha, Filipa de Castro, Carolina Campos, Paula Diogo, Catarina Miranda, Vera Santos, Xana Novais, Cristina Branco, Marta Cerqueira, Aldara Bizarro, Filipa Francisco, Ana Mesquita, Nélia Pinheiro, Rita Castro Neves, Cristina Peres, Vânia Doutel Vaz
Literatura	Matilde Campilho, Maria Velho da Costa, Agustina Bessa Luís, Dulce Maria Cardoso, Maria Gabriela Llansol, Anabela Mota Ribeiro, Alice Vieira, Bárbara Bulhosa, Maria Alzira Seixo, Hélia Correia, Rosário Pedreira, Alexandra Lucas, Clara Capitão, Carla Maia de Almeida, Inês Pedrosa, Maria Helena da Rocha Pereira, Clara Ferreira Alves, Zita Seabra, Guilhermina Gomes, Cecília Andrade, Natália Correia, Isabel Lucas, Luiza Neto Jorge, Maria de Lourdes, Teolinda Gersão, Teresa Calçada, Marta Martins, Isabel Alçada, Golgona Anghel, Maria Carlos Loureiro, Olga Ferreira Lima, Sílvia Alves, Cristina Taquelim, Maria Filomena Molder, Maria Filomena Mónica, Maria Cardoso, Adília Lopes, Paula Cristina Costa, Joana Matos Frias, Leonor Riscado, Maria Teresa Horta, Alice Munro, Sara Figueiredo Costa, Sandra Vieira Jürgens
Música	Ana Moura, Maria João, Carminho, Gisela João, Mariza, Joana Carneiro, Cristina Branco, Mísia, Amália Rodrigues, Selma Uamusse, Lula Pena, Rita Redshoes, Manuela de Freitas, Márcia, Aline Frazão, Raquel Tavares, Manuela Azevedo, Maria da Fé, Maria

João Pires, Susana Santos Silva, Katia Guerreiro, Teresa Salgueiro, Ana Quintans, Lura, Surma, Capicua, Ana Deus, Ana Bacalhau, Celeste Rodrigues, Lucília do Carmo, Sara Serpa, Joana Seara, Fernanda Maria, Fábria Rebordão, Eduarda Melo, Cuca Roseta, Cátia Moreso, Joana Gama, Carla Caramujo, Amélia Muge, Maria Luísa de Freitas, Susana Travassos, Maria Teresa de Noronha, Susana Gaspar, Elisabete Matos, Ângela Alves, Joana Sá, Barbara Tinoco, Sandra Martins, Dora Rodrigues, Júlia Reis, Rita Maria, Sónia Tavares, Ana Miró, Sandra Medeiros

Teatro Joana Craveiro, Mónica Calle, Beatriz Batarda, Sara Carinhas, Eunice Muñoz, Regina Guimarães, Sandra Faleiro, Aida Tavares, Raquel André, Fernanda Lapa, Maria João Luís, Rita Blanco, Silvina Pereira, Paula Diogo, Carla Bolito, Manuela Maria, Crista Alfaiate, Vânia Rodrigues, Cátia Terrinca, Ana Vitorino, Sofia Dinger, Rita Cabaço, Rita Cabaço, Maria Gil, Isabel Abreu, Inês de Medeiros, Manuela de Freitas, Catarina Barros, Sofia Marques, Rita Durão, Maria do Céu Ribeiro, Alexandra Moreira da Silva, Maria João, Sara Ribeiro, Rita Loureiro, Lúcia Sigalho, Elisabete Pinto, Magda Bizarro, Joana Bárcia, Paula Mora, Lúcia Maria, Isabel Alves Costa, Anabela Almeida, Paula Só, Marta Bernarde, Cristina Reis, Márcia Breia, Isabel Muñoz Cardoso, Céu Guerra, Patrícia Portela, Ana Cris, Maria João Vicente, Ana Brandão, Rita Ribeiro, Maria João Abreu, Zia Soares, Sílvia Filipe, Micaela Cardoso, Catarina Wallenstein, Graça Lobo, Lia Gama, Ana Luena, Susana Madeira, Rita Brütt, Rita Cruz, Ana Guiomar, Marta Freitas, Rosinda Costa, Julia Bernat, Isabel Teixeira, Ana Moreira, Emília Silvestre, Maria João Brilhante, Joana Brandão, Virgínia Vitorino, Isabel Fernandes, Susana Sá, Mónica Garnel, Maria Mascarenhas, Carolina Passos Sousa, Maria Rueff, Ana Brito e Cunha, Paula Garcia, Vanda R Rodrigues, Teresa Coutinho

Artes Paula Rego, Joana Vasconcelos, Ângela Ferreira, Lourdes de Castro, Ana Pinho, Graça Fonseca, Raquel Henriques da Silva, Carla Filipe, Fernanda Fragateiro, Ana Jotta, Maria Helena Vieira da Silva, Isabel Carlos, Ana Vieira, Filipa Oliveira, Josefa de Óbidos, Cristina Guerra, Ana Vidigal, Wasted Rita, Helena Almeida, Graça Brandão, Paula Varanda, Gabriela Albergaria, Catarina Alfaro, Pauliana Valente Pimentel, Manuela Marques, Dalila Rodrigues, Elisabete Caramelo, Mariana Silva, Dalila Gonçalves, Grada Kilomba, Patrícia Almeida, Sara Antónia Matos, Maria João Vilhena de Carvalho, Priscila Fernandes, Laura Castro, Seabra Carvalho, Maria Trabulo, Joana Escoval, Adriana Molder, Mónica de Miranda, Amanda Levet, Pauline Foessel, Ana Mesquita

A partir dessa informação e a disponível no HTML do público.pt e arquivo.pt, desenvolveu-se um *website* que nos convida a descobrir entre frases, imagens, links, menus as artistas portuguesas.

Homepage:

Existo

(latim *exsisto*, *-ere*) *verbo intransitivo*

1. Ter existência, viver, subsistir, ser.
2. Durar.
3. Estar vivo.

Se um ser de um mundo exterior olhar para a forma como comunicamos, para nos perceber melhor

Se olhar lançar um web crawler para analisar um jornal e fazer parsing das notícias

Se utilizar Inteligência Artificial para identificar o género dos habitantes do planeta

O que encontra ?

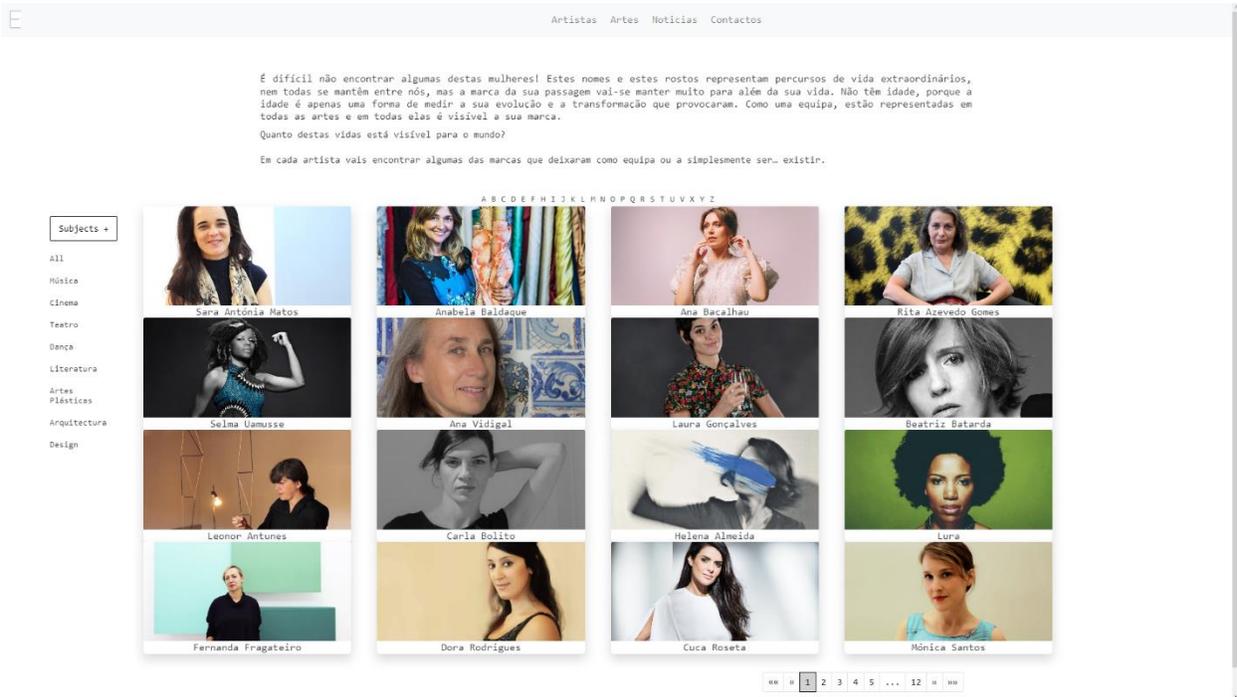
Será que nos encontra ? Será que existimos ?

Vem descobrir o que ele encontrou >>

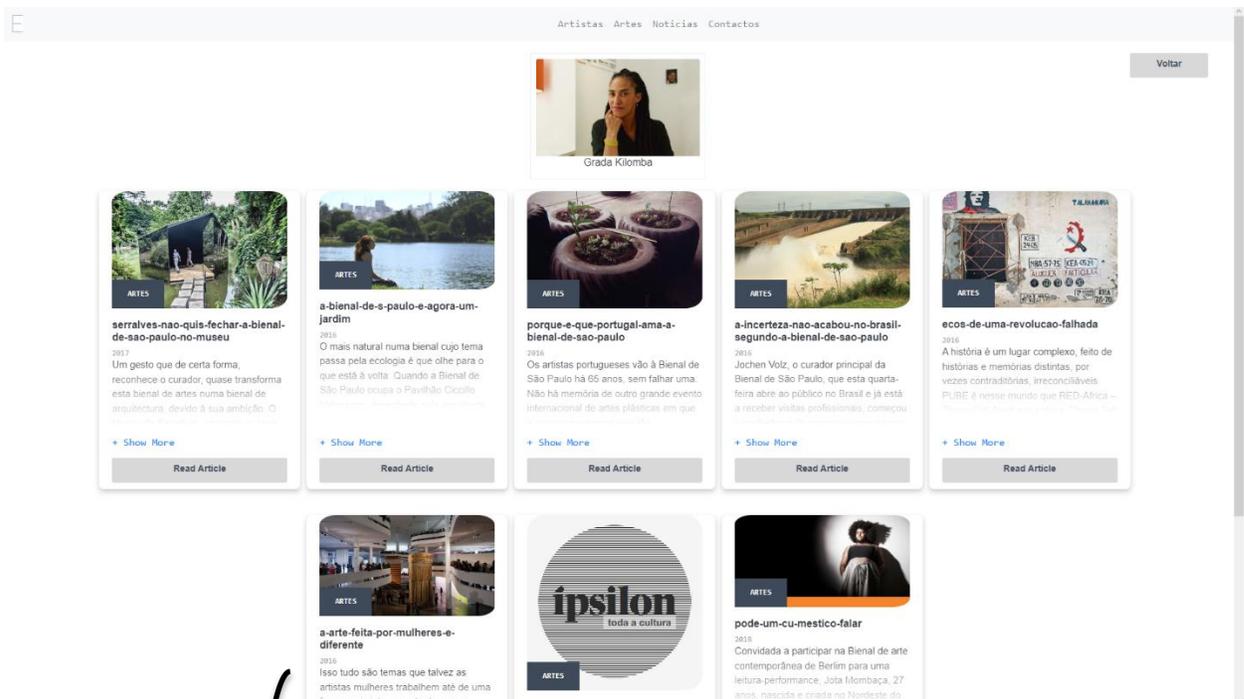


Esta página é o primeiro contacto que o utilizador têm com o website e pretende dar o contexto e suscitar curiosidade sobre a premissa do trabalho desenvolvido. Ao clicarmos na frase “Vem descobrir o que ele encontrou >>”, encontramos...

a página Artistas, onde o utilizador pode explorar e descobrir as várias artistas que encontramos que foram mencionadas em pelo menos 5 notícias. As artistas por ser ordenadas por ordem alfabética ou tipo de arte.



Cada artista tem associadas todas as notícias onde foi referida e quando se acede a uma dessas notícias, para além de ter acesso ao conteúdo fornecido pelo público (data, autores, sumário, etc), as frases onde a artista é citada ou referida aparecem automaticamente *highlighted*.



a-arte-feita-por-mulheres-e-diferente

Voltar

Isabel Salema | 08-09-2016 12:15:00

Keywords: B

Isso tudo são temas que talvez as artistas mulheres trabalhem até de uma forma mais interessante do que os artistas homens. Às três artistas mulheres portuguesas que estiveram em São Paulo perguntámos se a presença de uma maioria feminina faz da bienal uma exposição diferente. E atenção que a mulher negra ainda tem um status pouco respeitado, quantas mulheres artistas negras existem? Pessoalmente, Priscila Fernandes ainda se sente discriminada por ser mulher. Que mulheres, que homens? Para Grada Kilomba, "é absolutamente importante" esta ser uma bienal que teve mais mulheres do que homens presentes. Expandir o cânone? A presença de artistas mulheres tornou-se muito mais comum nas exposições.



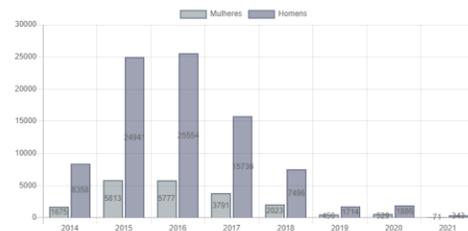
PubCom oito metros de altura, as torres de Luis Myrrha são a única obra verdadeiramente monumental desta Bienal de São Paulo, a grande exposição que durante três meses e até Dezembro ocupa o pavilhão Cicillio Matarazzo, desenhado por Oscar Niemeyer na cidade brasileira. A artista, uma mineira que vive em São Paulo nascida em 1974, quer mostrar nesta obra intitulada Dois Pesos, Duas Medidas (2016) como a arquitetura moderna brasileira não tomou em conta as culturas indígenas na sua maneira de construir, apesar de uma ou

A seguir somos convidados a explorar em forma de informação agregada e visualizada em *infographics* visões de uma mesma realidade: a presença ou visibilidade das mulheres e homens no mundo das artes, quem é que escreveu as notícias, e os pronomes mais usados.

Existir ... estar vivo. Mas também dar a conhecer aos outros que estamos vivos! Os gráficos desta página mostram diferentes visões de uma mesma realidade: a presença ou visibilidade das mulheres no mundo das artes, através das notícias. O universo das notícias é o das artes e a fonte o suplemento Ipsilon do jornal público. É o que esperavas? E o que se passa com as diferentes formas de arte? Será que elas existem?

A secção seguinte apresenta as notícias originais e está acessível através do menu "notícias" >>

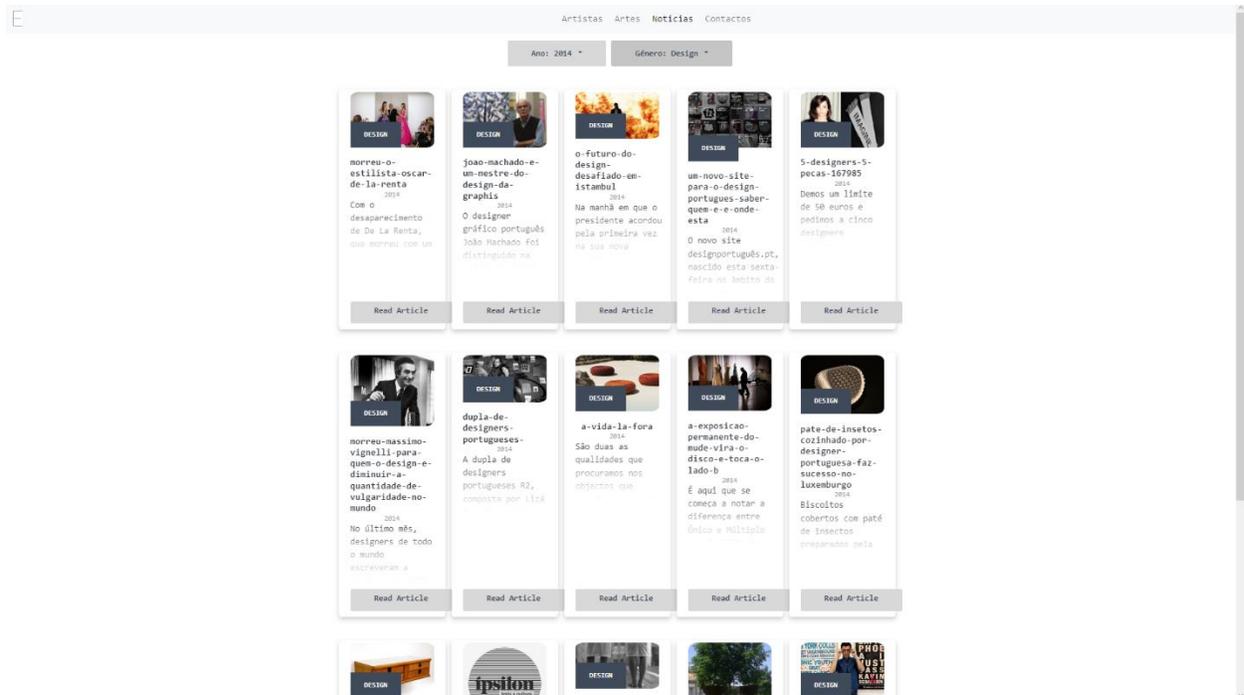
Música	Cinema	Teatro	Dança
Literatura	Artes Plásticas	Arquitectura	Design
All			



A partir de técnicas de processamento de língua natural, extraiu-se de forma automática as pessoas mencionadas no corpo da notícia. Representam as somas agregadas por género, as vezes que um artista ou uma artista são mencionados.

Notícias

Na página notícias estão disponíveis todas as notícias processadas e é possível filtrar por ano e rubrica.



Originalidade e carácter inovador

Este projeto é, tanto quanto nos foi possível observar, o primeiro a fazer uma análise desta temática em Portugal, pretendendo ser uma referência e um motor que promova o aumento da inclusão e visibilidade das mulheres artistas portuguesas.

Pretende-se apresentar a informação de uma forma objetiva, processada de forma automática, um instrumento de estudo para quem queira analisar de uma forma sistemática esta temática.

Impacto social (aplicação e utilidade social)

Este projeto permite de forma fácil e direta dar a conhecer as mulheres artistas e o seu trabalho a qualquer cidadão. Acreditamos que o seu impacto pode ser sentido a vários níveis:

- ilustrar a representatividade da mulher artista na imprensa, e forma como esta tem evoluído
- ilustrar as diferenças entre as várias artes, no que respeita à representatividade do género
- divulgar o trabalho das artistas representadas

Consequentemente poderá assim suscitar a curiosidade e o interesse pelas artes em geral.

Impacto científico (aplicação e utilidade científica)

A partir da informação disponibilizada, torna possível mais facilmente traçar o percurso de uma determinada artista, as pessoas com quem colabora, onde mostra o seu trabalho, e também a imagem/perspetiva que os críticos e outras pessoas na área da arte e do jornalismo tem sobre a mesma.

Relevância da utilização do Arquivo.pt

O projeto foi desenvolvido a partir de notícias arquivadas no Arquivo.pt, nomeadamente do suplemento Ípsilon do jornal Público. As *API's* disponíveis no *github* do Arquivo.pt foram usadas para obter todas as páginas de notícias recolhidos para alimentar o projeto. Sem o Arquivo.pt não teria sido possível realizar este projeto.

ⁱ <https://namsor.app/api-documentation.html#namsor-api>

ⁱⁱ <https://github.com/nicolasberube/Wiki-Gendersort>

ⁱⁱⁱ <https://pdfs.semanticscholar.org/de69/cad2e2ed9636739e106f76ec10ff441b4901.pdf>