

Premiada plataforma que recolhe dados sobre minorias em jornais

O Prémio Arquivo.pt distinguiu ainda uma ferramenta sobre as relações entre personalidades e um *site* que mostra a evolução de primeiras páginas de jornais

Teresa Sofia Serafim

Dois engenheiros informáticos criaram uma plataforma que permite recolher de jornais *online* dados sobre grupos minoritários, subrepresentados ou com algum estigma social. A ideia de Paulo Martins e Leandro Costa é conseguir mapear e estudar a representação desses grupos no contexto jornalístico português nas primeiras décadas do século XXI. Designado “Major Minors”, este foi também o projecto distinguido com o primeiro lugar, de 10.000 euros, pelo Prémio Arquivo.pt 2021 – um concurso que distingue projectos que mostrem a utilidade e importância do serviço Arquivo.pt, que preserva milhões de páginas recolhidas da *web* portuguesa.

Em segundo lugar (no valor de 3000 euros) ficou David Batista, que desenvolveu o projecto “Politiqices”. Esta ferramenta permite explorar e analisar relações de apoio e oposição entre personalidades em títulos de notícias. Já a terceira posição (no valor de 2000 euros) e a menção honrosa (uma assinatura digital de dois anos do PÚBLICO) foi para Susana Parreira. A designer criou uma plataforma – a “Primeiras páginas de jornais *online* portuguesas” – sobre o design gráfico de primeiras páginas de jornais *online* portugueses ao longo do tempo.

Esta é a quarta edição do Prémio Arquivo.pt 2021, um concurso promovido pela Fundação para a Ciência e a Tecnologia, do qual o PÚBLICO é parceiro mediático. Os trabalhos foram avaliados por um júri composto por Alfredo Ramalho (bibliotecário e arquivista), Celso Martinho (director de engenharia na empresa Cloudflare, Lisboa), Miguel Avillez (astrofísico da Universidade

de Évora) e Teresa Firmino (editora da secção de Ciência do PÚBLICO). Os prémios serão entregues a 30 de Junho, no encerramento do Encontro Ciência 2021, que terá a presença do Presidente da República.

Dos refugiados aos animais

O projecto “Major Minors” surgiu mesmo antes da possibilidade da candidatura ao Prémio Arquivo.pt. Paulo Martins conta que tudo começou com uma aliança entre o Departamento de Informática da Universidade do Minho – o berço do trabalho – e o Centro de Estudos Humanísticos da mesma instituição. “Os investigadores do Centro de Estudos Humanísticos precisam de dados e nós, do Departamento de Informática, também gostamos de dados e temos interesse em extraí-los e em trabalhar modelos para os tratarmos”, relembra o investigador da Universidade do Minho.

Com Leandro Costa, pós-graduado em engenharia informática na mesma universidade, desenvolveu uma base de dados através de recortes de imprensa de jornais portugueses que fizesses referência a questões e grupos minoritários, subrepresentados ou com algum estigma social. Os recortes pertencem ao PÚBLICO, porque é o jornal com uma representação mais antiga no Arquivo.pt. Foram recolhidos 16.000 artigos sobre esses grupos.

Para a recolha, desenvolveram um algoritmo que identifica automaticamente artigos, comentários e fotos que façam referência a oito grupos: refugiados, animais, asiáticos, africanos, homossexuais, migrantes, ciganos e mulheres. Além dos grupos, são abordadas questões de género ou raciais mais generalistas.

Esta plataforma está aberta a todos, mas poderá ser útil para gru-



Nesta página, David Batista; ao lado, Paulo Martins, Leandro Costa e Susana Parreira



O Centro de Estudos Humanísticos precisa de dados e o Departamento de Informática tem interesse em extraí-los

Paulo Martins

pos de investigação que estudem várias questões minoritárias – e não só. Agora, a intenção de Paulo Martins e Leandro Costa é convidar a comunidade científica a usar esta ferramenta. Para isso, vão apresentá-la na Slate 2021, um colóquio de ciências de informação e informática em Esposende, a 1 e 2 de Julho.

E já há planos para o futuro. “Futuramente, gostávamos de a expandir”, assinala Paulo Martins, que trabalha na área da linguística computacional. Por exemplo, gostariam de replicar o conceito para outras questões ou para outros modelos de dados. Há ainda o desejo de começar a extrair automaticamente artigos de outros jornais.

Relações de apoio e oposição

Ainda quando estava a fazer o seu doutoramento, David Batista tinha tido a ideia de fazer algo semelhante

ao “Politiqices”. Actualmente a trabalhar em processamento de linguagem natural na *startup* Comtravo, em Berlim, decidiu avançar: “A ideia tinha ficado sempre em águas de bacalhau, mas comecei a brincar com isto e pensei que o Arquivo.pt tinha imensos dados e que estava aberto o concurso.”

Tirou então a ideia da gaveta e desenvolveu uma ferramenta que permite explorar e analisar relações de apoio e oposição entre personalidades políticas em títulos de notícias. “Aqui em Berlim o Inverno é frio e escuro e, ainda para mais em tempos de covid-19, quando ficava de quarentena em casa, comecei a trabalhar nisto.”

Para este trabalho, David Batista recolheu artigos no Arquivo.pt de jornais *online*, como no *site* do PÚBLICO, da TSF, do *Expresso*, do *Diário de Notícias* ou da *Visão*. A par-



tir da Wikidata (uma espécie de Wikipédia para uma aplicação de *software*), reuniu nomes e informações de personalidades afiliadas a partidos políticos portugueses, com cargos públicos em Portugal ou pessoas que tiveram profissão de juiz ou economista. Na sua recolha na Wikidata acabou por ficar com uma base de dados de cerca de 1400 personalidades portuguesas relevantes.

“Depois conjuguei essas duas fontes de informação e extrai uma informação extra: sempre que há um tipo de notícia em que duas personalidades políticas são mencionadas, tento perceber se há uma interacção entre elas e se essa interacção se expressa numa relação de oposição, suporte ou em nenhuma delas”, explica.

Tendo tudo isto, ficou com uma colecção de dados que treinou com o recurso a um algoritmo de *machine learning* (aprendizagem automática) para que através de padrões nas palavras nos títulos a ferramenta pudesse concluir se havia as tais relações de apoio, oposição ou nenhuma delas.

Também encontrou limitações: como alguns títulos eram mais ambíguos ou irónicos, era mais difícil identificar uma relação de oposição ou apoio. Para isso, em futuros trabalhos, pensa mesmo estender a procura dessa relação ao texto do artigo.

Esta pode ser uma ferramenta de apoio à investigação jornalística. “Um jornalista pode querer ter uma colecção de notícias com relações de apoio ou oposição à volta de personalidades ou partidos”, exemplifica. Também pode ser usado para detectar padrões. Como exemplo, David Batista diz que Eduardo Ferro Rodrigues aparecia muito nas notícias num certo período, deixou de aparecer e depois reapareceu. “Está relacionado com o facto de se assumir alguns cargos de relevância.”

Futuramente, David Batista planeia “limpar um bocadinho mais o *site*”. E já recebeu sugestões de uma amiga que trabalha num jornal e está a usar a ferramenta em trabalhos para as autárquicas: porque não usar jornais locais? Algo que será desafiante, porque terá de alimentar essa base de dados, o que não é tão fácil como a que desenvolveu agora. Muitas vezes, não há muita informação sobre os presidentes de juntas de freguesias, como há de outros políticos a nível nacional.

Por fim, pretende ainda elaborar um artigo científico para investigadores na área da sociologia, informática ou ciência política sobre este trabalho.

A evolução dos jornais

Susana Parreira está bem consciente da enorme importância dos jornais e que a primeira página é usada para captar a atenção do leitor. Estes foram os pontos-chave que a levaram a desenvolver o projecto

“Primeiras páginas de jornais *online* portugueses”. Portanto, enquanto estava ainda a fazer a sua dissertação de mestrado na Faculdade de Ciências e Tecnologia da Universidade de Coimbra, desenvolveu esta ferramenta para conseguir perceber a evolução das primeiras páginas de jornais *online*.

Nessa ferramenta analisa-se o design gráfico e as funcionalidades de primeiras páginas de jornais *online* portugueses ao longo do tempo. Do PÚBLICO há 21 páginas de 21 anos diferentes (de 1998 a 2019, excepto de 2018). Há ainda algumas páginas do *Expresso* e do *O Jogo*. Através do *site* que Susana Parreira criou, pode-se assim mostrar a evolução do *design* dessas primeiras páginas dos jornais seleccionados.

“Depois de fazer a recolha e de ter diversos tipos de parâmetros – como a presença de imagens numa página ou a possibilidade de partilha de um artigo nas redes sociais –, transpuser esses dados para dois artefactos de visualização de informação para que facilitassem a visualização desses dados a qualquer pessoa”, explica Susana Parreira, que actualmente é designer na empresa tecnológica The Loop Company, em Coimbra.

Num primeiro artefacto (uma peça que ajuda a visualizar os dados recolhidos) converteu a primeira página num desenho mais simples – ficando apenas rectângulos a representar diversos parâmetros (cada um desses elementos é representado pela mesma cor).

Já um segundo artefacto permite visualizar a proporção que cada um dos parâmetros representa na página. Desta forma, conseguiu-se perceber qual o espaço ocupado por cada um deles – por exemplo, consegue-se ver qual o espaço ocupado pelas imagens e pelo texto.

Entre as observações que fez, Susana Parreira diz que é fácil ver que, ao longo dos anos e com a evolução tecnológica, o texto passou a ser menos e as imagens ganharam espaço. “Conseguimos perceber que a mancha que representa o texto vai sendo cada vez menos e a ocupação da imagem aumenta”, realça a designer, que no mestrado teve a orientação de Ana Sabino, Ana Boavida e Fernando Penousal Machado – professores da Universidade de Coimbra.

Esta ferramenta pode ser usada para investigação na área do design e do jornalismo, bem como no contexto da sala de aula. Pode também ser a base para projectos noutras áreas.

Por agora, o objectivo de Susana Parreira é manter o *site* o mais actualizado possível e ir acrescentando primeiras páginas de outros anos, assim como alargar a análise a outros jornais. Gostaria ainda de reunir informações e exemplo com as conclusões que esta nova ferramenta permite obter.

Se um artigo científico omite ratinhos no título, as notícias seguem-no

Teresa Sofia Serafim

Este é um dos primeiros estudos com provas científicas sobre a relação entre os títulos dos artigos científicos e os das notícias

Marcia Triunfol e Fábio Gouveia estão bem conscientes da complexa relação entre a forma como os cientistas divulgam as suas descobertas com ratinhos e a maneira como os jornalistas as noticiam. Por isso, exploraram o lugar dos ratinhos de laboratório nos títulos dos artigos científicos e nas notícias. Agora, num trabalho publicado na revista *PLOS Biology*, os dois investigadores sugerem que, quando os autores do artigo científico omitem logo no título que o estudo foi feito com ratinhos, os jornalistas que escrevem sobre essa investigação tendem a fazer o mesmo.

A investigação em doença de Alzheimer foi a área escolhida para desmitificar a hipótese de que a forma como os cientistas relatam as suas descobertas tem uma influência nas notícias. Afinal, muitas vezes, os títulos das notícias sobre o estudo desta doença não referem que os resultados foram obtidos através de ratinhos.

Analisaram-se assim 623 artigos científicos publicados em 2018 e 2019 que usaram ratinhos em estudos sobre Alzheimer. Para isso, dividiram-se os artigos em dois grupos: os que referiam nos títulos que os roedores foram o principal animal usado no trabalho; e os que omitiram essa informação no título. Depois, os investigadores analisaram se havia diferenças entre esses grupos e o que a comunicação social publicava.

Verificou-se que havia uma ligação entre os títulos dos artigos científicos e os das notícias. No geral, os jornalistas omitiam nos títulos as espécies estudadas se os cientistas também o tivessem feito. Mais: se os artigos científicos não mencionassem os ratinhos nos títulos, esses trabalhos recebiam uma maior cobertura mediática e havia mais *tweets* sobre o estudo.

“O facto de os cientistas omitirem que o estudo foi feito em ratinhos logo no título dos artigos acaba por levar a notícias em que também não colocam nos títulos que os estudos foram

feitos em ratinhos”, resume Marcia Triunfol, consultora científica da Sociedade Humana Internacional, com sede nos EUA. A investigadora que vive em Lisboa destaca que existem estudos que mostram que a maioria das pessoas só lê os títulos: “Se não se colocar logo no título que o estudo foi feito com roedores, pode ficar-se com a impressão de que a investigação se aplica também a pessoas.”

De acordo com os autores, com que seja do seu conhecimento, este é o primeiro estudo a apresentar provas científicas de que a forma como os cientistas divulgam os seus resultados tem um papel nas notícias. Mesmo assim, os dois investigadores destacam num comunicado sobre o trabalho que há algumas limitações no estudo. Por um lado, os artigos analisados incluem só publicações de acesso aberto. Por outro, apenas se analisaram estudos em Alzheimer.

Quanto a estudos futuros, Marcia Triunfol adianta que se quer investigar porque é que os cientistas escolhem omitir os ratinhos nos títulos dos seus trabalhos. No artigo na *PLOS Biology* já se levantam algumas hipóteses, mas a investigadora diz que não foi possível tirar conclusões: “Não será pela escolha da revista nem pelas limitações das instruções da revista. Como é feita essa decisão?”

Marcia Triunfol espera que nos próximos tempos comece a existir algum tipo de recomendação para que os resultados de estudos com animais tenham a sua menção no título. “É algo que esperamos que aconteça e que as principais revistas percebam essa necessidade”, nota.

Para quem quiser acompanhar esta questão, há uma conta no Twitter sobre ela: a @justsaysinmice. Criada por James Heathers (da Universidade do Noroeste, nos EUA), chama a atenção para publicações com títulos que omitem que os resultados foram conseguidos em ratinhos.

“Conseguimos perceber [nas primeiras páginas dos jornais *online* portugueses] que a mancha que representa o texto vai sendo cada vez menos e a ocupação da imagem aumenta

Susana Parreira

Muitos dos estudos sobre Alzheimer são feitos com ratinhos

