

# Prémio Arquivo.pt 2019

## Identificação

- Título: Revisionista
- Área temática: Análise de alterações em notícias ao longo do tempo (Tecnologia/Comunicação)
- Candidatos: Flávio Martins e André Mourão
- Email: [fnm@fct.unl.pt](mailto:fnm@fct.unl.pt), [a.mourao@campus.fct.unl.pt](mailto:a.mourao@campus.fct.unl.pt)
- Link para a plataforma desenvolvida: <https://revisionista.pt>

## Descrição do Trabalho

Os media têm o crucial papel de divulgar informação de confiança de forma atempada e transparente [1]. Cabe aos mesmos, como bastiões da verdade, combater a crescente desinformação que se acentua com o advento das redes sociais e da disseminação de "fake-news" que faz cada vez mais parte do quotidiano contemporâneo.

Longe vão os tempos em que o principal veículo noticioso eram as publicações com suporte em papel, montados de véspera. Neste suporte, os jornalistas e editores têm constrangimentos que afectam a correcção ou actualização das peças, pois estas não podem ser alterada após seguir para impressão. Segundo Maier [2] menos de 2% dos artigos de jornais com erros factuais são seguidos de uma correção.

O surgimento do suporte online trouxe mais agilidade e flexibilidade às peças jornalísticas que podem agora ser alteradas a qualquer momento. No entanto, a flexibilidade oferecida pelo online permite que os artigos sejam alterados sem qualquer aviso de uma forma pouco transparente, que pode gerar desconforto e críticas dos leitores quando descobertas [3]. Visto que assumir essas alterações fica ao critério da publicação, nada garante ao leitor que um artigo não foi alterado. Quando partilhamos um link para um artigo, podemos ter a certeza de que o conteúdo se manterá igual? E se foi alterado, será este facto claramente assinalado?

Para responder a estas questões, criámos uma extensão para browser, Revisionista X, que permite ao utilizador saber se o artigo que está a ler sofreu alterações e apresenta as mesmas integradas no texto do artigo.

De forma a termos uma visão mais global, apresentamos uma plataforma de análise global da imprensa portuguesa, com base no processamento de grandes quantidades de notícias dos jornais online. Nesta ferramenta, podemos ver o histórico de revisões de cada artigo, pesquisar no texto editado e analisar as estatísticas de como são as edições para os diferentes jornais.

## Motim em prisão brasileira de Manaus faz pelo menos 6055 mortos

Um ~~tumulto~~ ~~motim~~ que durou 17 horas ~~no interior de numa na maior~~ prisão de Manaus, capital do estado do Amazonas, no Brasil, provocou pelo menos 55 mortos. ~~Inicialmente o balanço oficial indicava pelo menos 60 mortos, noticiaram vários meios de comunicação social brasileiros. A Secretaria de Segurança Pública do estado teme que haja mais vítimas mortais, mortais, mas o número foi actualizado ao fim da tarde desta segunda-feira.~~ Houve detidos que aproveitaram a rebelião para fugir, alguns dos quais foram, entretanto, recapturados. Contudo, ainda não há números oficiais das fugas.

O ~~motim tomou conta~~ fugir do Complexo Penitenciário Anísio

Revisionista X

The screenshot shows the Revisionista.pt interface. At the top, there is a navigation bar with links: Primeira Página, Destaques, Folhear, Estatísticas, Sobre X, and a search box labeled 'Pesquisar texto ou URL'. Below the navigation bar are filters: 'Todos os Tipos de Edição', 'Todos os Jornais', 'Ordenar', and 'Descendente'. The main content area displays two news items. The first item is titled '100 casas para travar saída de moradores do centro histórico de Lisboa' from 'Jornal de Notícias' (29 de Dezembro de 2017, 19:36). It shows a revision history for '1 de Janeiro de 2018, 09:01' with a red bar indicating removal and a green bar indicating addition, and a 'Comparar' button. The second item is titled 'Homem barra-se em posto dos correios na Ucrânia e faz reféns. Cinco já foram libertados. - Observador' (30 de Dezembro de 2017, 19:04). It shows a revision history for '1 de Janeiro de 2018, 06:45' with a red bar indicating removal and a green bar indicating addition, and a 'Comparar' button. Below the second item, the original text is shown with red strikethroughs and green additions: 'Homem barra-se Ucrânia. Policia libertou reféns e deteve homem barricado em posto dos correios na Ucrânia e faz reféns. Cinco já foram libertados. - Observador'.

Revisionista.pt

## Objetivos

O Revisionista foi criado para responder às seguintes questões:

- Que notícias são editadas?
- O que foi adicionado e removido?
- Onde estas alterações se encontram?
- Que tipos de alterações são feitas?
- Estas alterações são explícitas no texto?
- Que jornais editam mais informação?

## Resultados Atingidos

O Revisionista X é uma ferramenta, compatível com os browsers Chrome e Firefox, que permite ao utilizador saber se o artigo que está a ler sofreu alterações. Em caso afirmativo, o utilizador poderá aceder a um visualizador "Modo Leitura" que lhe permite ver de forma clara e unificada as diferenças entre a versão original da página e a sua versão actual. Neste modo, apenas o conteúdo do artigo é apresentado, tendo sinalizado a vermelho o texto removido, a laranja o texto alterado e a verde o texto adicionado.

No Revisionista.pt, apresentamos uma plataforma de análise global da imprensa portuguesa, com base no processamento de grandes quantidades de notícias dos jornais online. Nesta ferramenta, podemos ver o histórico de revisões de cada artigo, desde a sua versão original até à versão mais recente. As diferenças entre todas as versões, podem ser visualizadas sobre a forma de "Track changes". O objetivo é mostrar como os artigos são

alterados ao longo do tempo e se estas alterações são transparentes para os leitores. Para além de permitir explorar vários exemplos de artigos com revisões, o Revisionista.pt oferece uma ferramenta de pesquisa avançada com a qual o utilizador pode descobrir artigos revistos que correspondem aos termos de pesquisa.

## Originalidade e carácter inovador

Tanto quanto sabemos, nenhuma plataforma online de notícias permite auditar o histórico das edições de um determinado artigo publicado. Isto significa que muitas vezes o leitor não é notificado acerca destas alterações não visíveis! Com este trabalho devolvemos o poder ao leitor, que a partir de agora poderá usar o Revisionista X, uma extensão para browser, como um sexto sentido. Quando o leitor visita um qualquer site do domínio .PT será informado se esse mesmo site foi alterado e poderá escolher ver as diferenças para a versão original do artigo no ‘Modo de leitura’ de diferenças de uma maneira inovadora e compacta.

Para além da análise por artigo, os resultados também podem ser apresentados de maneira agregada para os sítios de notícias (e.g. tipos de edições e palavras mais adicionadas e removidas). Para esse efeitos, criámos a plataforma Revisionista.pt on utilizamos também processamento de linguagem natural e estatísticas para categorizar os tipos de edições (e.g. “Em Direto” para artigos que cobrem eventos a ocorrer no momento, “Correção Não Visível” para artigos alterados, sem menção a correções ou atualizações no texto) e quais as palavras mais removidas e adicionadas para os principais jornais portugueses.

## Impacto social (aplicação e utilidade social)

Este trabalho tem um papel chave para todos os cidadãos, pois ao mostrar o que foi alterado entre revisões de artigos, é possível aumentar a transparência dos media. O nosso objectivo é tornar este processo mais transparentes e confiável para todos os cidadãos. Nesta secção pudemos ver alguns exemplos das edições encontradas:

Grande Oriente Lusitano - Eleições na maçonaria deverão ter segunda volta

Arquivada a 3 de Junho de 2017, 19:30 Diário de Notícias 1 revisão encontrada

Arquivada a 4 de Junho de 2017, 19:57 Correção Invisível Comparar

Fernando Lima, atual Grão-Mestre, lidera com 41%	Fernando Lima obteve 37%
o advogado, quando só faltam votar 30 das 100 loja do GOL, já recolheu 41%	José Adelino Maltez 33% e
-Segue-se o economista	30%
31%, sendo que o professor universitário José Adelino Maltez apenas recolheu 28% dos votos dos mestres	

Neste exemplo, podemos ver a alteração do resultados das eleições na maçonaria enquanto os resultados ainda estavam a ser contados.

CDS-PP prefere eleições autárquicas a 1 de outubro – Observador

Arquivada a 27 de Março de 2017, 17:25 Observador 1 revisão encontrada

Arquivada a 29 de Março de 2017, 01:16 Adição Invisível Comparar

~~CDS-PP prefere~~ PSD, CDS, Verdes e PAN preferem eleições autárquicas a 1 de outubro – Observador

<del>CDS-PP prefere</del>	<a href="#">PSD, CDS, Verdes e PAN preferem</a>
<del>CDS-PP</del>	<a href="#">CDS</a>
<del>como "a melhor data"</del>	<a href="#">o dia</a>
<del>Para a líder do CDS-PP</del>	<a href="#">PSD</a>
<del>"dia 24 ainda é tempo de férias para muitas famílias"</del>	<a href="#">Verdes e PAN defenderam uma posição idêntica. O Governo regista "consenso". PS, BE e PCP serão ouvidos na quarta-feira</a>
<del>Autor</del>	<a href="#">Autores</a>
<del>=</del>	<a href="#">Observador</a>
<del>.</del>	<a href="#">O PSD, Os Verdes e o PAN também defenderam a mesma posição. No final dos encontros com o primeiro-ministro para a marcação das eleições, também o Governo veio registar esse "consenso" e dizer que a data final deverá ser conhecida na próxima quinta-feira, depois de decisão do Conselho de Ministros. "Tem havido um consenso sobre o dia 1 de outubro", disse a secretária de</a>

Neste exemplo, vimos a adição de partidos a uma posição inicialmente defendida pelo CDS-PP. Embora a adição seja legítima, não encontramos referências à edição na nova página.

2018 chega com aumentos nas portagens e nos transportes - Economia - RTP Notícias

Arquivada a 29 de Dezembro de 2017, 19:55 RTP Notícias 1 revisão encontrada

Arquivada a 31 de Dezembro de 2017, 09:11 Atualização Assumida Comparar

<del>..A título de exemplo, a Infraestruturas de Portugal assinala que "os preços das taxas de portagem da A21 (Ericeira-Venda do Pinheiro) não irão sofrer qualquer alteração em 2018"</del>	<a href="#">/ atualizado em 29 Dez, 2017, 20:43</a>
---	---

Neste exemplo, vemos que um parágrafo que referia que não iriam aumentar as portagens na A21 foi removido da notícia, tendo sido assumida no texto esta atualização. Decidimos investigar e encontrámos a fonte para esta afirmação no site das Infraestruturas de Portugal<sup>1</sup>

	Lanço	Taxas 2017			
		Cl 1	Cl 2	Cl 3	Cl 4
1	Ericeira - Mafra Oeste	0,5	0,85	1,1	1,2
2	Mafra Oeste - Mafra Este	0,65	1,1	1,4	1,6
3	Mafra Este - Malveira	0,5	0,85	1,1	1,25
4	Malveira - Venda do Pinheiro	0,3	0,55	0,7	0,8

Auto Estrada	Lanço	Extensão c/ Port. (km)	Calc. Previo 2018			
			Cl 1	Cl 2	Cl 3	Cl 4
A21	1 Ericeira - Mafra Oeste	5,2	0,55	0,95	1,20	1,35
	2 Mafra Oeste - Mafra Este	6,8	0,70	1,30	1,65	1,80
	3 Mafra Este - Malveira	5,3	0,55	1,00	1,30	1,40
	4 Malveira - Venda do Pinheiro	3,4	0,30	0,55	0,70	0,80
	total	20,7	2,10	3,80	4,85	5,35

Fomos verificar os preços das portagens em 2017<sup>2</sup> e 2018<sup>3</sup>, e confirmámos mesmo que este aumento existiu. Este é um caso de uma correção factual correta, que acabou por não ser corrigida numa potencial fonte oficial.

## Impacto científico (aplicação e utilidade científica)

Relativamente ao impacto científico, o Revisionista tem potencial científico em várias frentes:

- Plataforma: plataforma pesquisável de edições de notícias;
- Dados: dataset de diferenças de notícias portuguesas presente na plataforma;
- Ferramentas: as ferramentas desenvolvidas permitem a investigadores fazer novos trabalhos de investigação sobre notícias.

A plataforma online permite pesquisa avançada e disponibiliza vários filtros de dados, permitindo assim investigar para qualquer tema as modificações ao longo do tempo. Assim, a sua utilização permite a cientistas de áreas tão diversas como as ciências sociais, económicas e políticas investigar e reportar sobre todos os temas, sem a necessidade de recolher e processar as páginas das notícias.

As ferramentas desenvolvidas podem ser aplicadas a um conjunto mais alargado de sites portugueses, internacionais e tirar partido de colaborações com outros arquivos da Internet. Investigadores que desejem expandir este trabalho ou aplicar o mesmo tipo de técnicas a outros domínios ou países, poderão fazê-lo usando as ferramentas desenvolvidas:

- O bulk\_downloader permite descobrir e descarregar as notícias que foram editadas para um domínio para um certo período de tempo;

<sup>1</sup> [“Atualizações dos Preços de Portagem 2018”](#), IMTT, Visitado a 2 de Maio de 2019

<sup>2</sup> [“Taxas de Portagem A21”](#), IMTT, Arquivado a 18 de dezembro de 2017

<sup>3</sup> [“Taxas de Portagem A21”](#), IMTT, Arquivado a 29 de abril de 2018

- O `html_cleaner` elimina todo o conteúdo não relevante da notícia, baseado num conjunto de tags e outras anotações dadas (domínios conhecidos), ou através de técnicas heurísticas (novos domínios);
- A nosso analisador de diferenças no texto permite uma melhor representação das diferenças apresentando as palavras alteradas (ao invés de linhas ou caracteres).

Para dar continuidade a este trabalho com melhorias e funcionalidades adicionais, prevemos a submissão de um artigo científico para a conferência Web Search and Data Mining (WSDM), com uma análise aprofundada dos resultados obtidos para um número alargado de domínios. Este artigo irá demonstrar a nossa plataforma, os resultados obtidos e seu impacto.

## Relevância da utilização do Arquivo.pt

Este trabalho não seria possível sem o Arquivo.pt, pois foi utilizado como a fonte de todos os dados recolhidos. Para realizar este trabalho, foram recolhidas 139561 artigos do Arquivo.pt, dos quais 6793 artigos foram editadas mais do que uma vez. Estes artigos correspondem a doze sítios de notícias (Correio da Manhã, Diário de Notícias, DNotícias, Jornal Económico, Jornal de Negócios, Jornal de Notícias, Observador, Público, RTP Notícias, SIC Notícias, Sol, Visão). Um dos objetivos futuros é colaborar com o Arquivo.pt para aumentarmos a rapidez desta recolha, de forma a que seja possível analisar grandes volumes de dados de um número mais alargado de fontes noticiosas.

## Comentários adicionais

Durante a criação desta ferramenta, deparámo-nos com os mais variados tipos de informação. Um exemplo em particular mostrou-nos os desafios relacionados com a arquivação de informação: uma referência a um nome de uma testemunha que foi substituído por um nome fictício numa revisão da notícia. Devido à sensibilidade desta informação, decidimos apagar essa edição da nossa base de dados e enviar um pedido ao Arquivo.pt para eliminar a versão inicial da página com o nome verdadeiro. A Internet também traz um novo desafio para os jornais; a competição de outros sites, blogs e plataformas de social media, que produzem conteúdo gratuito e que não têm de respeitar as regras e código deontológico inerente a profissão do jornalista. Esta competição leva a uma diminuição das receitas e despedimentos, especialmente nos jornais impressos [4].

Para além da competição, a Internet também traz novas oportunidades para monetizar conteúdo. Jornais portugueses estão a começar a implementar modelos de subscrição que limitam o número de artigos para não subscritores, ou criam artigos de acesso exclusivo a assinantes [5] [6]. Embora este tipos de soluções sejam legítimas, estas reduzem a capacidade do Arquivo.pt de recolher informação histórica relevante. No futuro, gostaríamos que o Revisionista trabalhasse em conjunto com jornais e o Arquivo.pt para garantir que toda a informação é preservada de uma maneira que não ponha em causa a monetização (e.g. respeitando o período de embargo, e retrações críticas como a mostrada no exemplo anterior do nome fictício).

## Recursos complementares

- [1] [“Competir com as “fake news” será a morte do jornalismo, alerta António Costa”](#) - Rádio Renascença, 25 de abril de 2019
- [2] Maier, S. , [“Tip of the Iceberg: Published Corrections Represent Less than Two Percent of Factual Errors in Newspapers”](#), “2007 Annual meeting of the Association for Education in Journalism and Mass Communication”, The Renaissance, Washington, DC
- [3] [“Why Did the New York Times Change Their Brooklyn Bridge Arrests Story?”](#) - The Village Voice, 02 de outubro de 2011
- [4] [“Newsroom employment dropped nearly a quarter in less than 10 years, with greatest decline at newspapers”](#), Pew Research Center, 30 de julho 2018.
- [5] [“O acesso à edição digital do PÚBLICO vai mudar. Saiba como e porquê”](#), PÚBLICO, 7 de abril de 2019
- [6] [“Premium - Observador”](#), Observador, Visitado a 2 de maio de 2019