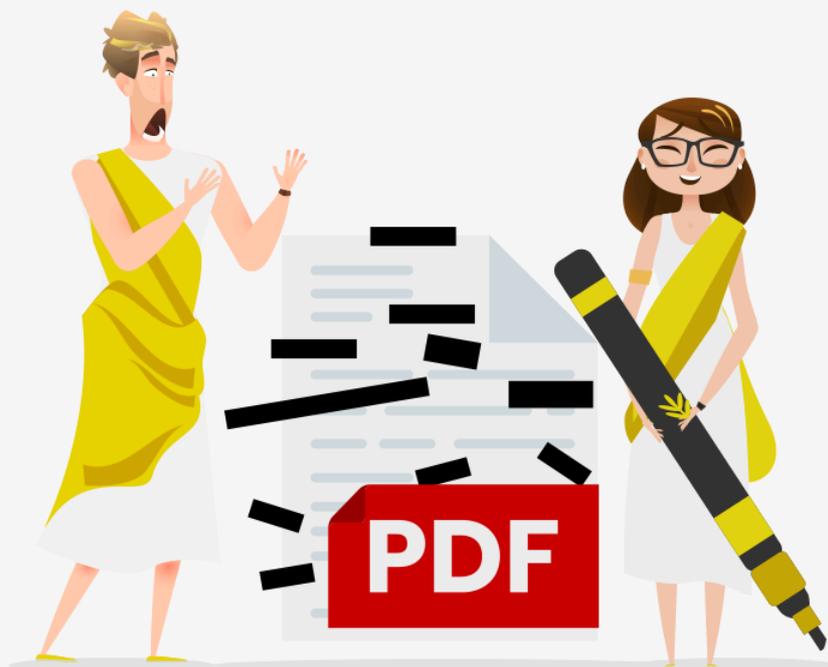


Falhas Épicas De Redação De PDF, Uma História De Terror



Leia este artigo em francês (</fr/redaction-pdf-erreurs-a-ne-pas-reproduire/>).

Muitas ferramentas estão disponíveis para redigir documentos eletrônicos, mas nem todo mundo está familiarizado com elas. O post de hoje mostrará os erros de redação que viraram notícia nos últimos anos. Não reproduza

Índice



- O que é redação?
 - Etimologia
 - Da redação à censura
 - Do tribunal...
 - O relatório Mueller
 - ...para o escritório
- Erros épicos de redação de PDF: 3 exemplos
 - O Facebook gosta dos seus dados pessoais
 - Paul Manafort e a tecnologia
 - O contrato da AstraZeneca com a UE
- Como evitar falhas de redação em PDF
 - Copiar e colar e outras dicas
 - Mas o trabalho não acabou
 - Sanitização
 - Uma palavra sobre metadados
- Resumindo
 - Experimente nossa ferramenta de redação online

O que é redação?

Em sua definição mais neutra, redação é o processo de remoção permanente de elementos visíveis de um documento.

Já escrevemos sobre redação em arquivos PDF em dois artigos anteriores:

- Como redigir um PDF online e remover dados pessoais e confidenciais (<https://blog.avepdf.com/three-free-tools-secure-pdf-files/>)
- Três ferramentas gratuitas para proteger seus arquivos PDF (<https://blog.avepdf.com/three-free-tools-secure-pdf-files/>)

No post de hoje, mostraremos os erros que você não deve reproduzir ao redigir um PDF. Mas antes, vamos contar um pouco da história dessa técnica.

Etimologia

De acordo com Merriam-Webster (<https://www.merriam-webster.com/news-trend-watch/barr-redacted-report-coming-in-april-20190329>), o verbo redact é usado em inglês desde o século XV, derivando do latim redigere “dirigir, liderar ou trazer de volta, reunir, coletar, organizar, reduzir”.

No início do século XIX, redação significa simplesmente organizar ou editar (https://www.cjr.org/language_corner/redacted-censored-mueller-report.php). Ainda encontramos esse significado em francês, onde “un rédacteur” é um editor, geralmente em uma mídia.

Da redação à censura

Além de sua definição neutra baseada na etimologia, o ato de redação muitas vezes assume um aspecto moral com a noção de censura.

Desde que a escrita existe, tentamos esconder palavras sensíveis.

Recentemente e graças às novas tecnologias de raios-x, as cartas de Maria Antonieta ao Conde de Fersen revelaram seus segredos. Embora o conteúdo de sua correspondência seja principalmente político, as frases em que Fersen escreve sobre seus sentimentos são redigidas (por ele mesmo!) para evitar a censura.

bensaabe — raine

bensaabe — raine

Os segredos das cartas de Maria Antonieta ao conde de Fersen revelados. Arch. nat., 440AP/1. © CRC

A linha entre redação e censura (https://www.cjr.org/language_corner/redacted-censored-mueller-report.php) é tênue.

Em inglês, uma definição de redação (https://www.americanbar.org/groups/judicial/publications/judges_journal/2019/spring/embarassment-redaction-failures/) é “texto e imagens confidenciais em um documento que foram censurados, excluídos ou obscurecidos”.

Em francês, “biffer” e “caviarder” significam ambos redigir, mas o último também significa censurar. O primeiro significado de “caviarder” é, na verdade, “cobrir com uma camada preta (tão preta quanto caviar), tornar ilegível uma passagem de um texto proibido pelo censor”.

Dependendo do contexto em que o documento é produzido, pode-se considerar a redação um ato de censura. Felizmente, hoje em dia, falamos principalmente sobre redação em um contexto mais neutro, menos político e moral, como veremos a seguir.

Do tribunal...

Se questões legais impulsionaram o uso da redação em todos os países, nos EUA o aumento de sua adoção está ligado a dois eventos específicos.

Primeiro foi a necessidade de proteger testemunhas de organizações criminosas como a máfia quando elas eram mencionadas em documentos produzidos para tribunais. Então, desde o Freedom of Information Act (FOIA) ([https://en.wikipedia.org/wiki/Freedom_of_Information_Act_\(United_States\)](https://en.wikipedia.org/wiki/Freedom_of_Information_Act_(United_States))) em 1966, agências governamentais precisam disponibilizar uma grande quantidade de documentos ao público. No entanto, a divulgação desse tipo de documento só é possível se todos os elementos referentes a questões de segurança nacional ou informações pessoais e sensíveis forem removidos. Isso é possível com a "seção de isenção" do FOIA. E é por isso que muitos documentos que se enquadram no domínio da FOIA se parecem com isto:

~~CONFIDENTIAL~~

1 - Mr. A. Rosen
1 - Mr. T. J. Bishop

The Acting Attorney General April 25, 1972

Director, FBI 1 - Mr. E. S. Miller
1 - Mr. E. L. Shackelford
1 - Mr. T. J. Smith (Morner)
1 - Mr. E. L. Pence

JOHN WINSTON LENNON
SECURITY MATTER - NEW LEFT 1678 REF 16/4
4-5-72

9863 ADD/TS
7/2/93
CAF 83-1240
1678 REF/KG 1-16-86

John Winston Lennon is a British citizen and former member of the Beatles singing group.

(b1) b2

Despite his apparent ineligibility for a United States visa due to a conviction in London in 1968 for possession of dangerous drugs, Lennon obtained a visa and entered the United States in 1971. During [redacted]

(C)

Lennon had contributed \$75,000 to a newly organized New Left group formed to disrupt the Republican National Convention. The visas of Lennon and his wife, Yoko Ono, expired on February 29, 1972, and since that time Immigration and Naturalization Service (INS) has been attempting to deport them. During the Lennons' most recent deportation hearing at INS, New York, New York, on April 18, 1972, their attorney stated that Lennon felt he was being deported due to his outspoken remarks concerning United States policy in Southeast Asia. The attorney requested a delay in order that character witnesses could testify for Lennon, and he then read into the court record that Lennon had been appointed to the President's Council for Drug Abuse (National Commission on Marijuana and Drug Abuse) and to the faculty of New York University, New York, New York.

(b7D) 5
DATE 7/10/93
CLASSIFIED BY [redacted]

MAILED 2
APR 26 1972

100-469910
MLP:plm
(10)

CONFIDENTIAL
Group 1
Excluded from automatic
downgrading and
declassification

SEE NOTE PAGE TWO

ALL INFORMATION CONTAINED
HEREIN IS UNCLASSIFIED EXCEPT
WHERE SHOWN OTHERWISE

7/10/93
1972

HQ-11

No início dos anos 1970, o governo dos EUA conduziu vigilância sobre o ex-Beatle John Lennon. Esta é uma carta do diretor do FBI J. Edgar Hoover ao Procurador-Geral. J. Edgar Hoover, domínio público, via Wikimedia Commons

Não tínhamos documentos eletrônicos nem redação em PDF nos anos setenta, então ainda tínhamos que contar com um bom marcador preto para apagar informações.

Mas isso não significa que estamos melhores hoje, mesmo com as ferramentas certas amplamente disponíveis.

O relatório Mueller

O caso que tornou a redação de PDF famosa entre o público.

Tendências: 'redact', 'redacted'

Pesquisas aumentaram 4.000% em 29 de março de 2019

Merriam-Webster (<https://www.merriam-webster.com/news-trend-watch/barr-redacted-report-coming-in-april-20190329>)

Em 2019, o Departamento de Justiça dos EUA divulgou o relatório Mueller.

Para contextualizar, o promotor especial Robert Mueller foi o encarregado de investigar suspeitas de conluio entre Moscou e a equipe de campanha de Donald Trump em 2016. Em seu relatório, Mueller exclui qualquer conluio entre as equipes de Donald Trump e o poder russo. No entanto, ele destaca, sem poder decidir, uma série de incidentes que tendem a mostrar que o presidente buscou obstruir o andamento da investigação e questionamentos sobre possível obstrução do curso da justiça.

Usando a Lei de Liberdade de Informação, vários meios de comunicação, como o BuzzFeed News (<https://www.buzzfeednews.com/article/jasonleopold/robert-mueller-declined-charge-trump-jr-roger-stone>) e, mais tarde, a CNN, processaram para obter acesso ao relatório.

O Departamento de Justiça concedeu o desejo deles, porém com pequenas modificações.

De fato, cerca de um oitavo das linhas são redigidas, caindo na seção Exemption.

II. Russian “Active Measures” Social Media Campaign

22 pages



(<https://www.npr.org/2019/04/18/713974980/see-how-much-of-the-mueller-report-is-redacted?t=1649177391182>)

Uma amostra do relatório Mueller, conforme mostrado no artigo da NPR de 18 de abril de 2019.

Como nota lateral, há muitas coisas a dizer sobre a versão que o Departamento de Justiça forneceu ao público. Entre outros, os principais problemas foram:

- o relatório não era pesquisável,
- seu tamanho era muito, muito grande.

Ambos os aspectos tornaram o relatório difícil de compartilhar e pesquisar, o que não deveria ter sido assim, especialmente com um arquivo PDF.

Leia mais sobre como o DoJ poderia ter se saído melhor com este arquivo no artigo da PDF Association Uma avaliação técnica e cultural do Relatório Mueller em PDF

(<https://www.pdfa.org/a-technical-and-cultural-assessment-of-the-mueller-report-pdf/>) .

...para o escritório

Além do sigilo de defesa, todas as empresas e organizações no mundo todo precisam gerenciar dados pessoais e confidenciais.

Alguns setores têm a obrigação legal de não tornar públicos alguns de seus dados, como o setor jurídico com sigilo judicial e o setor da saúde com sigilo médico e profissional.

Mas mesmo para todas as outras empresas ou organizações, existem regulamentações em vários níveis para proteger funcionários e clientes.

- GDPR na Europa e, em particular, a CNIL na França;
- A Lei de Privacidade do Consumidor da Califórnia nos EUA;
- O Brasil implementou recentemente uma lei semelhante, e outros países estão seguindo o exemplo.

Os departamentos de justiça agem relativamente rápido neste tipo de caso. Por exemplo, oito meses após a implementação do GDPR, a CNIL na França reivindicou ao Google mais de 50 milhões de euros, e a multa foi confirmada pelo Conselho de Estado em 2020.

Em termos de sanções, as violações do GDPR podem custar até 20 milhões de euros ou 4% do faturamento global das empresas.

Erros épicos de redação de PDF: 3 exemplos

Infelizmente, há muitos casos de redação que deram errado.

No entanto, podemos agradecer às pessoas e empresas nos três exemplos a seguir por mostrar a todos as práticas a serem evitadas.



O Facebook gosta dos seus dados pessoais

Mesmo que as pessoas não confiem mais na famosa rede social (<https://www.zdnet.com/article/new-poll-shows-facebooks-severe-trust-problem/>), na verdade foi uma falha de redação que a impediu de vender dados pessoais a terceiros em 2012.

Em 2017, um PDF mal redigido mostrou que o Facebook considerou cobrar de grandes empresas pelo menos US\$ 250.000 para acessar os dados pessoais dos usuários. (<https://arstechnica.com/tech-policy/2018/11/facebook-pondered-for-a-time-selling-access-to-user-data/>)

O repórter da Ars Technica Cyrus Farivar descobriu o problema graças a uma simples cópia e colagem de um documento judicial de 2017 em um editor de texto como o Notepad ou um documento do Office. Este é o teste mais básico se você quiser saber se um documento está redigido adequadamente. Se você vir o conteúdo redigido no editor de texto, isso significa que o texto foi simplesmente ocultado e não removido.

Você verá no exemplo a seguir que esse também é o erro mais frequente que as pessoas cometem ao redigir um documento. Não é porque um texto está coberto de preto (com um marcador ou anotação) que ele não está mais lá.



Paul Manafort e a tecnologia

Paul Manafort foi presidente da campanha presidencial de Donald Trump em 2016.

Sua série de erros de manuseio de tecnologia é tão longa (<https://www.wired.com/story/paul-manafort-bad-tech-pdfs-passwords/>) que faz dele um exemplo clássico para campanhas de segurança digital (pense em reutilizar senhas e e-mails antigos, deixar de converter documentos e armazenar mensagens incriminatórias na nuvem).

Sua equipe de advogados não sabia melhor quando tentou redigir uma passagem sensível. Infelizmente, simplesmente copiando e colando o parágrafo redigido, foi possível ler as partes apagadas e descobrir novos detalhes sobre o relacionamento de Manafort com Konstantin Kilimnik (https://www.vice.com/en_us/article/8xpye3/paul-manafort-russia-case-redaction-fail), um antigo associado com laços com a Rússia.



O contrato da AstraZeneca com a UE

Parece que podemos ver uma *leve* melhoria tecnológica em nosso próximo caso.

No início do ano passado, e por questões de transparência, a Comissão Europeia publicou o contrato redigido que assinou com a empresa de vacinas AstraZeneca (<https://euobserver.com/health-and-society/150799>).

À primeira vista, o contrato parece redigido corretamente com as ferramentas adequadas. No entanto, os marcadores do PDF referentes ao conteúdo redigido foram ignorados. (<https://www.pdfa.org/pdf-redaction-astrazeneca-eu-contracts-st-happens/>)

Sobre o que era a parte redigida? O preço total de compra deste contrato (€870m).

Como evitar falhas de redação em PDF

A tecnologia de exclusão permanente de dados em PDFs não é nova, mas (aparentemente) ainda não é muito conhecida.

Na verdade, já em 1998 a empresa Appligent registra uma patente sobre edição de PDF (como um lembrete, a Adobe lançou a primeira versão do PDF em 1993). Em 2006, na versão 1.7, a Adobe adicionou um adendo sobre anotações de redação.

Copiar e colar e outras dicas

Então vimos o que acontece quando você simplesmente coloca retângulos pretos ou anotações em informações confidenciais graças ao teste de copiar e colar; você não apaga o texto ou a imagem abaixo, apenas os oculta.

Também tomamos nota de verificar nossos favoritos graças à AstraZeneca, mas você também precisa verificar outros lugares, como:

- legendas de imagens
- hiperlinks
- arquivos incorporados

- anexos

Mas o trabalho não acabou

Como regra geral, com PDF, você deve sempre se perguntar: “se eu não vejo, significa que não está lá?”

Como qualquer outro documento eletrônico, o PDF é composto de uma parte visível, o que vemos quando o abrimos, e uma parte invisível, que fica mais ou menos escondida.

O problema com informações sensíveis? Você pode encontrá-las nas partes menos visíveis do PDF. Então, vamos procurá-las.

Sanitização

Se a redação é o processo de remover informações visíveis de um documento eletrônico, a higienização é o mesmo processo, mas para informações ocultas (ou prefiro a expressão “menos visíveis”).

Por informações ocultas, queremos dizer:

- metadados do arquivo (autor, título, data de criação, versão em PDF, etc.),
- arquivos e metadados de imagens incorporados,
- anotações,
- comentários,
- camadas de texto ocultas (camada OCR, abaixo das anotações).

Frequentemente, dados sensíveis que você cuidadosamente gastou tempo identificando em um PDF podem aparecer em outro lugar. Por exemplo, você removeu o nome de uma pessoa do texto, mas ele também pode ser encontrado nos metadados de imagens ou em comentários.

Embora a higienização, assim como a redação, possa ser automatizada, muitas vezes é necessário fazer uma verificação manual, especialmente nos seguintes casos:

- imagens em diferentes formatos,
- planilhas,
- arquivos anexados,
- índices.

Uma palavra sobre metadados

Os mecanismos de busca podem acessar metadados, portanto, tenha muito cuidado ao publicar um documento redigido online.

Metadados podem conter virtualmente qualquer informação e podem ser relacionados ao documento em geral ou a objetos separados contidos no documento, como imagens, fontes, etc.

O software de leitura padrão geralmente não acessa todas essas informações, mas ainda pode ser extraído por soluções avançadas de processamento de PDF.

Algo que os desenvolvedores devem ter em mente ao redigir documentos.

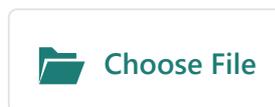
Resumindo

Com a redação de PDF, usar as ferramentas certas é essencial. Mas o que é crucial é verificar todas as fontes de informação.

No caso de processos complexos de redação semântica que envolvem mais do que simplesmente remover números de telefone ou de previdência social (processos simples que podem ser facilmente automatizados), nenhuma ferramenta pode vencer uma pessoa treinada.

Experimente nossa ferramenta de redação online

Drop your PDF file here or



Max file size: 128 MB

Se você estiver procurando por um SDK de redação avançado, confira nossos kits de ferramentas para desenvolvedores (<https://www.orpalis.com/products/#toolkits>) .