CAPTCHA

Pedro Anjos(45558), Rafael Silva(45813)

Maio 2021

1 Abstract

Se é um utilizador assíduo da 'internet', com certeza que já viu caracteres ondulados e verificações de que não é um robô. Mas alguma vez se questionou o porquê de os resolver, e qual a sua importância na 'internet'? Esta investigação tem por base responder a essa questão, bem como realçar o seu papel na 'internet'.

2 Tópicos

2.1 CAPTCHA

O termo "CAPTCHA"foi pela primeira vez introduzido em 2000 por Luis von Ahn, Manuel Blum, Nicholas Hopper and John Langford na Carnegie Mellon University. Um CAPTCHA consiste em num desafio visual, na forma de caracteres alfanuméricos distorcidos, ou auditivo, e visa assegurar que a entidade que os resolve é um humano[11].

A razão da sua eficácia relaciona-se com a complexidade para com os "softwares robóticos", grau de dificuldade esse não sentido pelo homem que os resolve facilmente.

Por exemplo, quando vamos comprar bilhetes para um determinado espetáculo na internet, somos submetidos à prova de um CAPTCHA, com o objetivo do próprio site prevenir os comportamentos suspeitos dos scalpers.

Não só prevê as atuações dos scalpers, com também são muito eficazes contra os spammers, sendo muito frequentes em sites onde são possíveis interações entre utilizadores, ficando suscetíveis a ataques automáticos de bots.

Por exemplo, blogs desprotegidos são alvos destes ataques feitos pelos spammers, que pretendem enviar sucessivos comentários automáticos que anunciam, de forma indesejada pelo blogger, os seus produtos. Uma possível solução é o uso dos CAPTCHA, que limita a escrita de comentários apenas para os humanos.

Além da sua eficácia, são estimados que existam cerca de 200 Milhões de CAPTCHA, sendo facilmente gerados em múltiplos sites.

2.2 TIPOS DE CAPTCHA

Gerar um CAPTCHA significa apresentar um desafio e dar a hipótese de resposta a um humano ou software robot. Os mesmos são classificados de acordo com tipo de problema[10].

1. CAPTCHA's baseados em textos;
2. CAPTCHA's baseados em imagens;
3. CAPTCHA's baseados em áudio;
4. CAPTCHA's baseados em vídeos

1) São muito fáceis de implementar, sendo eles muito eficazes e com uma grande amplitude de questões, desde decifrar caracteres corretos, até resolver uma simples questão.

Figura 1. CAPTCHA's baseados em textos

2) São testes em que o utilizador é posto à prova com imagens para descobrir qual se encontra na posição correta, resolver puzzles, ou as imagens que descrevem uma determinada palavra.

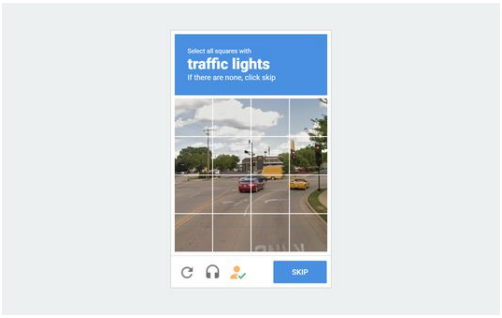

Figura 2. CAPTCHA's baseados em imagens

3) Este tipo foi desenvolvido para usuários com dificuldades visuais, impossibilitados de resolver os tipos 1. e 2. Contém um áudio distorcido criado pelo programa que escolhe uma sequência de dígitos e palavras aleatórios. O clip é então apresentado ao usuário terá de o escutar e produzir manualmente o que ouviu.

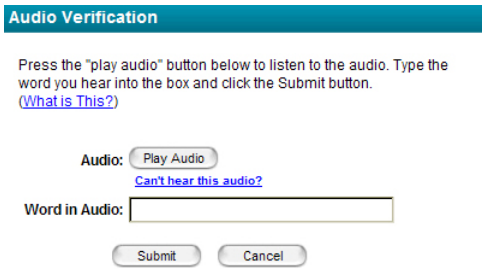

Figura 3. CAPTCHA's baseados em áudio

4) É proposto ao utilizador que apresente três palavras (tags) que descrevam corretamente o vídeo apresentado. Caso a tag descreva corretamente o vídeo, o desafio é superado.

A necessidade de evoluir os testes CAPTCHA e criar provas cada vez mais complexas provém do avanço das tecnologias.

Figure 3: Video based CAPTCHA

Figura 4. CAPTCHA's baseados em vídeos

2.3 reCAPTCHA

Em 2007, a Google criou um serviço CAPTCHA gratuito intitulado de re-CAPTCHA. A ideia base seria idêntica, mas começar a ter em atenção o tempo necessário para resolver uma prova. De acordo com entidades da Google, em 2010 estimava-se a existência de 200 milhões de diferentes CAPTCHA's. Contabilizando 10 segundo para a resolução de apenas um, se os decifrássemos todos seriam necessárias 500 mil horas.

O ponto de partida desta nova implementação seria pegar num livro e digitalizar um determinado excerto do mesmo e pôr um computador à prova para o decifrar através da tecnologia OCR (Optical Character Recognition) [12]. Por defeito, esta tecnologia não reconhece certos caracteres, comprometendo a sua resolução. O objetivo seria tirar proveito da falha desta tecnologia, aproveitando as palavras irreconhecíveis pelo OCR, e dispor um teste CAPTCHA capaz de as decifrar corretamente. A dúvida é, como é possível apresentar uma resposta para uma palavra que o sistema de verificação não o reconhece? A solução seria gerar 2 palavras num único CAPTCHA, em que uma delas, que nós não sabemos qual é, é reconhecida pelo computador, e a outra não. Caso o utilizador decifre a palavra conhecida por parte do computador, o mesmo assume que o utilizador é um humano. [5]

A Google criou também a famosa seleção da certificação "Eu não sou um robô". Este tipo de teste não se trata de apreender um bot, mas sim evitar fraudes e manter a segurança para os utilizadores num determinado site. Por exemplo, um software desenvolvido para invadir uma conta no Gmail está programado para colocar infinitas senhas até acertar[3]. E como deteta um comportamento fraudulento? A Google utiliza duas vias para detetar o ataque dos mesmos, sendo eles: [2]

1) Verificar o comportamento da verificação. O ser humano usa um rato de forma irregular e imperfeito. Já nos robôs, os mesmos estão programados para usar o rato do ponto A até B. Este tipo de comportamento é detetado e o seu acesso é bloqueado;

2) Ter em conta os Cookies de um utilizador na internet. Quando navegamos na internet e pretendemos aceder a um site, somos como obrigados a aceitar Cookies. Quanto mais assídua for a nossa presença na rede, maior é o número de Cookies, à priori não somos robôs;

Por fim, existem imensos sites que recorrem ao reCAPTCHA, pois possuem um equilíbrio entre a usabilidade e a segurança. Os testes são facilmente resolvidos pelo ser humano, e dispõe uma hierarquia de níveis de dificuldade. O desafio é maior consoante o número de CAPTCHA's resolvidos num curto espaço de tempo.

2.4 DESVANTAGENS

Apesar dos avanços da tecnologia contribuírem diretamente para uma maior complexidade dos testes CAPTCHA, os hackers prosseguiram o mesmo rumo, sendo por isso impossível de possuir uma rede 100% segura. Um bom exemplo para confirmar esta afirmação provém do recurso base do reCAPTCHA, a tecnologia OCR, com um dicionário cada vez mais amplo, capaz de reconhecer inúmeros caracteres e palavras. Por isso mesmo, os sites estão mais expostos ao perigo e sofrem sucessivos ataques. Segundo um artigo, um programador Russo invadiu os mecanismos CAPTCHA do site Yahoo, com 35% de sucesso[10]. Também os CAPTCHA do Microsoft live mail são sucessivamente alvos de ataques de spam. Além do mais, como os sites não têm acesso específico aos CAPTCHA que foram contornados, serão superados infinitas vezes.

Uma possível solução a este problema seria implementar reCAPTCHA com base nos testes auditivos, sendo esta uma prática pouco usual e recorrente em determinadas ocasiões ou sites.

Ademais, se estes desafios tiverem erros, o utilizador nunca consegue comprovar a sua identidade, levando à sua frustração e retirada do site. Exemplo disso são os testes para selecionar imagens onde aparece um determinado objeto. Certamente já o levou à dúvida uma imagem com uma parte microscópica desse mesmo objeto, e questiona se o deve selecionar ou não. Se isso for desprezado pelo criador, vai errar o teste e terá de fazer outro. E se o seguinte possuir o mesmo defeito? Leva à desistência pela perda de tempo, conceito colmatado pela Google[7].

Também nos testes reCAPTCHA, existem grandes discrepâncias de dificuldade dos níveis mais baixos para os mais extremos, e por vezes nos patamares mais elevados o próprio humano não é capaz de os resolver. Segundo a Forbes, em 2018 a Baymard Institute estimou que 8% dos usuários errava nos CAPTCHA de texto, subindo essa percentagem para 29% perante case-sensitive[1].

Por fim, na resposta ao reCAPTCHA que envolve a questão "Eu não sou um robô", surge o paradoxo da invasão de privacidade. Apesar de a navegação pela internet ser feita de forma livre, ter acesso a todos os nossos cookies significa ter conhecimento de todos os sites por onde pesquisa. Não que seja algo prejudicial para o utilizador, mas é de certa forma dispor, imensas informações pessoas à Google.

Contudo, esse problema já foi denunciado por várias empresas e defensores de privacidade e o site já se encontra a desenvolver um plano para acabar com os cookies de terceiros e criar uma nova tecnologia, intitulada de FLoC (Federated Learning of Cohorts). O objetivo é agregar utilizadores em grupos conforme as suas preferências, por exemplo, indivíduos que gostam de música clássica e râguebi[8]. Porém, esta solução equaciona várias dúvidas por parte de empresas, que receiam que a Google obrigue o uso de novas ferramentas, que aumentem substancialmente as receitas económicas do site.

2.5 ECONOMIA

Em 2020, com o aparecimento da Covid-19, o CAPTCHA e o reCAPTCHA tiveram um papel fundamental no que diz respeito à economia. Devido às restrições impostas pela pandemia, seria mais importante que nunca possuir uma rede segura onde os utilizadores pudessem realizar as suas aquisições em segurança[4].

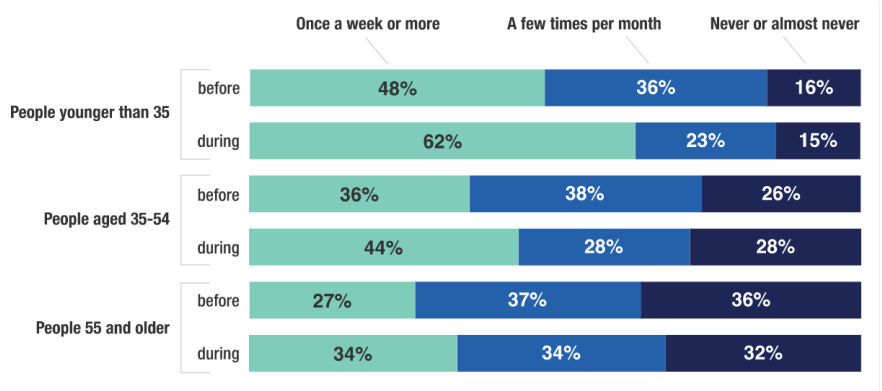

Figura 5. Impacto da Pandemia no comércio online[4].

Contudo, as limitações provocadas pela pandemia influenciaram a ação dos hackers, que por diversas vezes contornavam os CAPTCHA para tirar proveito das suas intenções. Por exemplo, na época mais lucrativa do ano, o Natal, a Playstation lançou a PS5, sendo apenas possível adquiri-la de forma online. Essa

restrição levou a que indivíduos criassem softwares para comprar imensos exemplares assim que o produto estivesse disponível. O objetivo não seria consegui-lo por interesse próprio, mas sim revendê-lo pelo dobro ou triplo a fanáticos da consola, tendo também a seu favor as unidades limitadas pela Sony. Processo esse que resultou, pois, durante e logo após a época natalícia esta tecnologia só se encontrava disponível a vendedores particulares a preços superiores aos 500 euros[9].

Uma curiosidade acerca este tópico, existem certas pessoas que pagam a indivíduos pela resolução de testes CAPTCHA. Com o passar dos anos, houve uma diminuição no valor da recompensa, pois em 2007 seriam pagos por cada 1000 CAPTCHA's 10 dólares, já em 2009 só era pago 0,75 dólares por cada 1000 CAPTCHA's resolvido[6].

2.6 PROPOSTA DE INOVAÇÃO

Revendo o subtema 2.4, com a implementação da tecnologia FLoC, a Google deixará de ter acesso a certos tipos de dados e hábitos de pesquisa de cada utilizador, o que de certa forma irá condicionar diretamente a tecnologia CAPTCHA, que tira proveito da utilização dos cookies. Isto equaciona a questão sobre o seu funcionamento futuro: teremos o regresso dos exaustivos testes para provar que não somos robôs? Para propor uma possível resolução, teremos como base não só o tempo investido na resolução dos testes, como também o que diferencia entre o ser humano e um robô.

Ao contrário dos autómatos, o Homem possui de uma anatomia própria com capacidades físicas e motoras que o distinguem dos outros seres vivos. A tecnologia vem, por exemplo, auxiliar na monitorização da saúde de um indivíduo, como, por exemplo, medição dos níveis de oxigénio no sangue, qualidade do sono, medição do número de batimentos cardíacos por minuto, entre outros. Um exemplo de uma dessas tecnologias é o Smartwatch.

Com o aparecimento deste engenho, o relógio deixou de ser um gerador de código de tempos permanente e passou a oferecer ao utilizador imensos serviços, entre os quais a monitorização de saúde. Atualmente, muitos destes dispositivos são criados por uma empresa de Smartphone, que já inclui no seu sistema uma aplicação capaz de sincronizar com o Smartwatch e uma conta de endereço eletrónico, para assim ser possível visualizar e armazenar todos os dados.

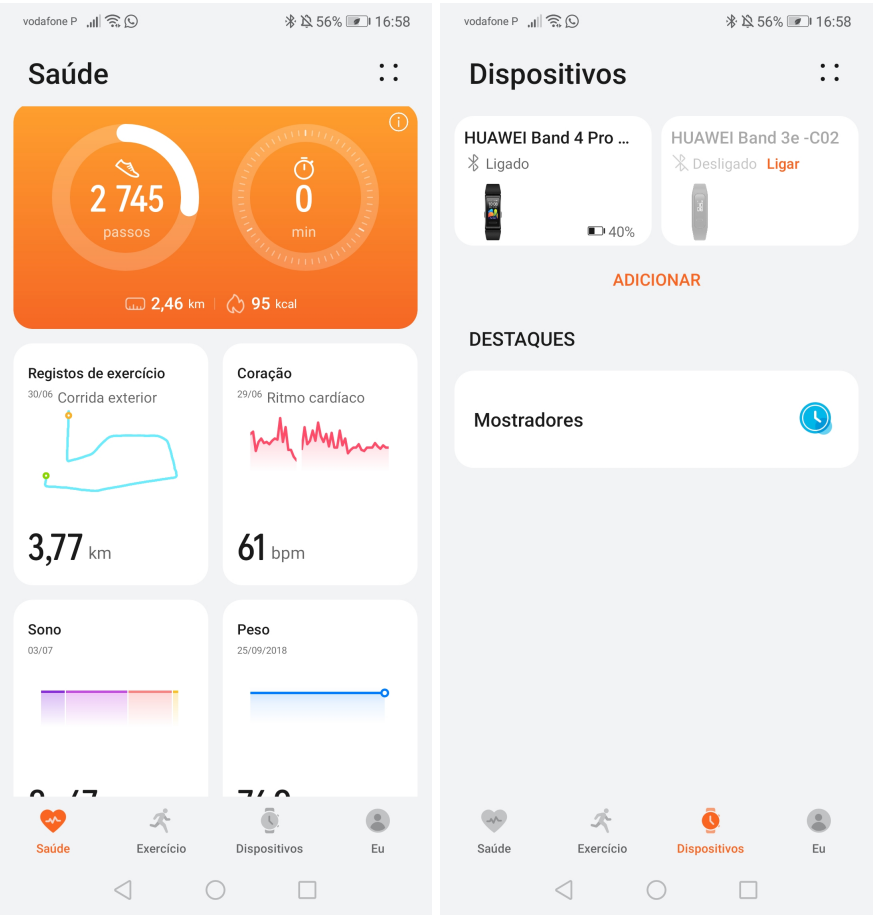

Figura 6. Printscreen da aplicação "Health"Huawei

O acesso aos dados de saúde por parte dos Smartwatch é também utilizado para auxiliar a prática de exercício físico, bem como melhorar a performance. Por exemplo, um indivíduo que corra com este aparelho, no final do seu treino tem a seu dispor na aplicação dados importantes acerca da sua corrida, tais como o efeito do treino aeróbico, o volume de oxigénio máximo, o ritmo cardíaco médio, entre outros. Além da corrida, a caminhada, a natação, o ciclismo, o remo ou o treino ao ar livre são atividades onde se podem obter este tipo de dados.

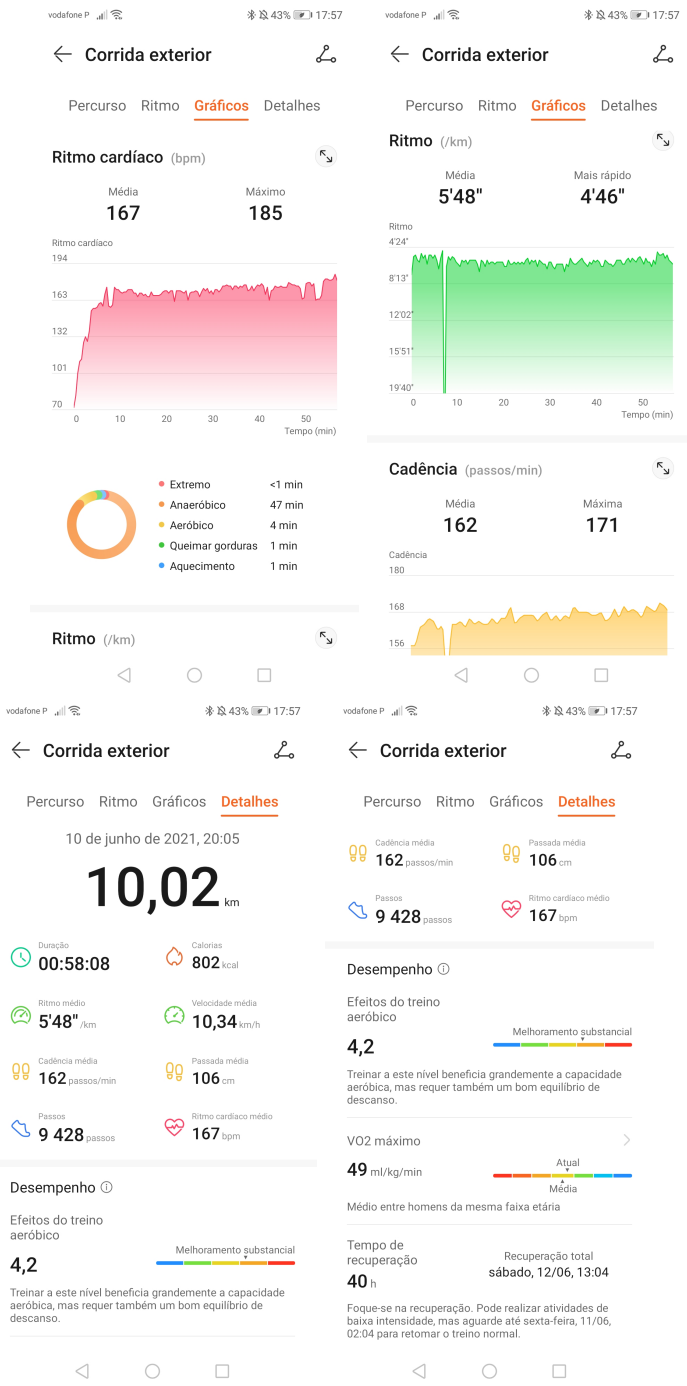

Figura 7. Disposição dos dados obtidos por um smartwatch Huawei, na aplicação Health

E como o CAPTCHA pode tirar proveito desta tecnologia? Ao iniciar a sessão num motor de busca com a mesma conta de correio eletrónico associada à aplicação, o site teria conhecimento de todos os dados de saúde e atividade física armazenadas na conta, provando que o utilizador não é um robô, e já não seria necessário verificar a sua identidade. O controlo de fraude seria feito na própria aplicação, com uma amostra de dados possíveis num ser humano, e em caso de irregularidade nesses dados, a conta seria bloqueada.

E para os indivíduos que não usem a aplicação, ou que não possuam um smartwatch, ou mesmo não pratiquem exercício físico? Os mesmos seriam submetidos a provas CAPTCHA para verificar a sua identidade, mas apresentando testes que promovam a atividade física. O objetivo mantinha-se, reduzir o tráfego de rede dos robôs, mas em simultâneo, incentivar o utilizador à prática de exercício físico.

Uma condicionante deste fundamento é o uso de um Smartwatch: estarão as pessoas interessadas em investir o seu dinheiro nesta tecnologia? Atualmente, existem várias gamas de aparelhos que monitorizam os dados apresentados anteriormente, com um custo mais baixo a rondar os 30 e os 50 euros. Porém, com o avanço tecnológico estão a ser criados aparelhos cada vez mais sofisticados, desvalorizando um pouco os atuais. Os menos valorizados, que já estariam fora do stock, podiam fazer parte de um kit, na compra do smartphone, para servir de amostra do produto.

3 CONCLUSÃO

A utilização desta tecnologia é uma mais-valia para colmatar os sucessivos ataques dos hackers e spammers, filtrando a rede dos mesmos. Com o avanço das tecnologias, este ramo tem evoluído de forma a evitar os contornos dos robôs e diminuir o seu tempo de resolução para proveito do homem. Contudo, apesar de ser impossível existir uma rede sem bots, a complexidade dos CAPTCHA tem influenciado na ação do ser humano, que cada vez mais erra e desiste mesmo antes de entrar no site. De certa forma até podemos afirmar serem dois opostos que permanecem de mão dada. Em relação à nossa proposta, e com recurso aos Smartwatch, o objetivo dos novos testes CAPTCHA não seria obrigar as pessoas à prática de exercício físico, mas sim incentivá-las a sair da inércia e se exercitarem.

Referências

[1] Captchas have an 8% failure rate, and 29% if case sensitive. `https://baymard.com/blog/captchas-in-checkout,urldate=14/05/2021`.

[2] Como os captchas funcionam — o que quer dizer captcha? `https://www.cloudflare.com/pt-br/learning/bots/how-captchas-work/,urldate=13/05/2021`.

[3] Surge in phishing attacks using legitimate recaptcha walls. https://www.helpnetsecurity.com/2020/05/01/recaptcha-walls/ ,urldate=15/05/2021.

[4] Liisa Ecola Charlene Rohr, Hui Lu. How is covid-19 changing americans' online shopping habits?, 2020. https://www.rand.org/pubs/research_reports/RRA308-6.html,urldate=17/05/2021.

[5] Jake Kenny. Google's recaptcha defeated by security researchers. https://www.kaspersky.com/blog/googles-recaptcha-defeated-by-security-researchers/11880/ ,urldate=17/05/2021.

[6] Chris Kanich Damon Mccoy Marti Motoyama, Kirill Levchenko. Captcha-solving services in an economic context, 2010. https://www.researchgate.net/publication/221260545_Re_CAPTCHAs-Understanding_CAPTCHA-solving_services_in_an_economic_context,urldate=14/05/2021.

[7] Chuck Pearson. Why your captcha is killing conversions, 2016. https://medium.com/rareview/why-your-captcha-is-killing-conversions-f9be6fe17d1f,urldate=14/05/2021.

[8] Karla Pequenino. Google avança com solução para eliminar cookies. o plano é "esconder indivíduos na multidão", 2021. https://www.publico.pt/2021/01/25/tecnologia/noticia/google-avanca-solucao-eliminar-cookies-plano-esconder-individuos-multidao-1947888, urldate=13/05/2021,.

[9] Ido Safruti. Your captcha could be hurting your sales, 2019. https://www.forbes.com/sites/forbestechcouncil/2019/08/07/your-captcha-could-be-hurting-your-sales/?sh=2f76058d33c8, urldate=16/05/2021.

[10] Baljit singh Saini. A review of bot protection using captcha for web security, 2013. https://www.researchgate.net/publication/272719923_A_Review_of_Bot_Protection_using_CAPTCHA_for_Web_Security, urldate=14/05/2021.

[11] Monica ChewJ. D. Tygar. Image recognition captchas. https://link.springer.com/chapter/10.1007/978-3-540-30144-8_23, urldate=14/05/2021.

[12] Michael Wyszomierski. Protect your site from spammers with recaptcha, 2010. https://developers.google.com/search/blog/2010/01/protect-your-site-from-spammers-with,urldate=17/05/2021.