Centro Federal de Educação Tecnológica de Minas Gerais Coletor de propósito geral

André Borges Yulli Dias

Introdução

O coletor é um *software* capaz de acessar os *links* da *web* e coletar os *htmls* das páginas. Neste trabalho, desenvolvemos um coletor de propósito geral, descrito https://andrelbol.github.io/SimpleCrawler/, que acessa *links* encontrados a partir de *urls* sementes. Essas sementes estavam pré-definidas na proposta do trabalho, para o André foi definida a terra.com.br, a Silvia escolheu uma das sementes que não estava sendo usada por nenhum aluno da turma e a Yulli havia ficado com a semente https://www.telegrafi.com/. Entretanto, tivemos o erro 403 Forbidden ao acessar essa url, por algum motivo o site não permite o acesso por robôs, então escolhemos http://www.zerozero.pt como a semente.

Utilizamos a arquitetura proposta no trabalho, representada na figura 1, onde a classe *Page Fetcher* tinha como responsabilidade descobrir novas *urls* que seriam gerenciadas pelo escalonador, responsável por oferecer as próximas páginas a serem visitadas pelo *Page Fetcher*. Além disso o escalonador tinha outras responsabilidades, dentre elas: verificar se a profundidade de páginas estava sendo respeitada, se os servidores estavam sendo acessados com uma janela mínima de 30 segundos e controlar a fila de páginas.

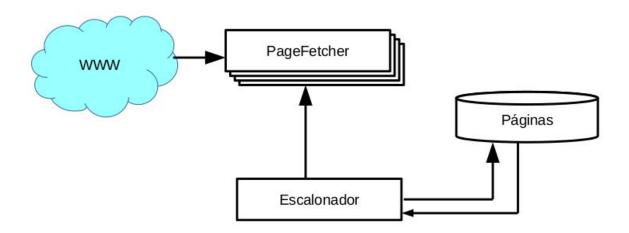


Figura 1 - Imagem da arquitetura. Fonte: Slides professor.

Como desafios encontrados neste trabalho podemos citar a dificuldade encontrada para acessar páginas mal-formadas. A utilização do *Html Cleaner* trouxe uma dificuldade para a coleta adequada de links, além disso tivemos que tratar *links* que não eram páginas válidas. Além disso, ao executar o projeto com uma número de threads maior que 30, algumas threads tiveram um estouro de pilha.

Para avaliar o número adequado de threads a serem utilizadas, executamos o programa variando a quantidade de threads de 10 a 100 de 20 em 20, conforme mostrado na tabela 1. De acordo com essa análise, escolhemos utilizar 10 threads nesse trabalho, pois com esse número conseguimos uma maior velocidade de coleta.

Número de Threads	Número de páginas Coletadas (N)	Tempo total de Coleta (s) (T)	Velocidade (N/T)
10	504	436	1,15596
30	524	562	0,93238
50	544	490	1,11020
70	564	525	1,07428
90	584	554	1,05415
100	595	583	1,02058

Tabela 1 - Velocidade de acordo com o número de Threads

Desenvolvimento

O desenvolvimento do coletor foi feito na linguagem Java, possuindo uma estrutura formada por 6 classes essenciais.

URLAddress

Classe escolhida como principal do projeto. Possui em sua estrutura os atributos:

- address: endereço pertencente à url a ser coletada
- depth: profundidade dessa url em relação à coleta sendo realizada

Essa classe permitia a formatação das *urls* buscadas, carregando consigo o metadado referente à quantidade de *urls* "mães" necessárias até encontrar a mesma.

Como decisão de projeto, essa classe foi escolhida também para abrigar o método **main** do coletor, sendo assim o local onde as *threads* de busca são iniciadas com as sementes de coleta.

Servidor

Classe responsável por armazenar informações dos servidores onde a coleta é realizada. Possui os atributos:

- **nome**: Armazena o nome do servidor
- **lastAccess**: Armazena um *timestamp* do último acesso realizado ao servidor, sendo esse no formato de milissegundos.
- ACESSO_MILIS: Constante que indica o tempo de intervalo mínimo entre acessos a esse servidor

A partir das informações de instâncias dessa classe, a coleta pode persistir as informações dos servidores por onde passou, além de conseguir respeitar o critério de tempo entre

acessos a um mesmo servidor (tempo indicado pela constante, escolhido como **30 segundos**). A classe ainda possui uma implementação da função **hashCode**, a qual permite a identificação única de cada servidor armazenado, evitando possíveis repetições indesejadas.

ColetorUtil

Classe de utilidades, possuindo um conjunto de métodos que auxiliaram na estruturação da coleta. Os métodos existentes são:

- **getUrlStream** e **consumeStream**: Métodos auxiliares na busca das páginas html referentes às *urls* coletadas, além de leitura do conteúdo das mesmas incluindo bufferização e decodificação
- **isAbsoluteUrl**: Método de verificação da existência de protocolo na *url* passada como parâmetro.

Record

Representação do robots.txt recuperado dos servidores buscados. A partir da utilização de métodos da biblioteca <u>irobotx</u> da empresa <u>TrigonicSolutions</u>, as instâncias dessa classe permitem a interpretação das regras explicitadas por cada servidor, facilitando as verificações de permissões de coleta. O método **allows** percorre todas as regras encontradas e verifica se o caminho passado como parâmetro a ser coletado viola alguma, sendo assim o principal método dessa classe.

EscalonadorSimples

Classe responsável por possuir as *urls* a serem coletadas, além de coordenar a busca para que seja uma busca em largura através das inserções e buscas personalizadas. Os atributos da classe são:

- **PROFUNDIDADE_MAXIMA**: Constante contendo a maior profundidade de coleta permitida às *threads*, servido como um ponto de parada da execução
- MAXIMO_DE_PAGINAS: Constante contendo o número limite de páginas a serem buscadas, servindo como um ponto de parada da execução
- numeroDePaginasColetadas: Variável controladora contendo o número de páginas coletadas até então
- filaDePaginas: "fila" de páginas armazena um mapeamento de servidores e suas urls buscadas. Sempre que uma nova url é encontrada na coleta e respeita condições específicas, é adicionada a esse mapeamento com seu respectivo servidor.
- **ur1sConhecidas**: Variável auxiliar do sistema que possui todas as urls que já foram percorridas. Auxilia no impedimento de repetições na coleta.
- listaDeRecords: Mapeamento entre servidores e seus respectivos <u>Record</u>s. A
 partir desse mapeamento é possível resgatar as configurações de coleta do servidor
 específico, necessário ao decidir se será adicionada a *url* à fila de páginas.
- urlsColetadas: Lista de urls que já foram coletadas.
- **horaDeInicio**: Variável que contém o *timestamp* de início da coleta (em milissegundos).
- numDeThreads: Variável que contém o número de threads utilizadas na busca

Para cada coleta realizada existe somente um escalonador, de forma que todas as *threads* de coleta compartilham esse escalonador.

Contendo e disponibilizando as urls de busca, essa classe coordena toda a coleta, afinal define o que cada *thread* irá realizar. Isso é realizado a partir dos seguintes métodos:

- **getUr1**: Método que retorna a próxima *url* na fila que esteja **disponível**. Essa disponibilidade é verificada mediante ao intervalo de tempo entre coletas no servidor referente à *url*. Caso não exista nenhuma url disponível, o escalonador ordena a *thread* a aguardar 1 segundo e percorrer novamente a lista de *urls*.
- adicionaNovaPagina: Método que adiciona uma nova url à fila de páginas, adicionando um novo servidor caso a url pertença a um servidor ainda não pesquisado.
- **getRecordAllowRobots**: Retorna o <u>Record</u> referente ao servidor da *url* passada como parâmetro.
- **putRecorded**: Adiciona um novo **Record** referente a um servidor passado como parâmetro.
- finalizouColeta: Verifica se a coleta chegou ao fim mediante à variável de controle MAXIMO_DE_PAGINAS, ou seja, se o número de páginas coletadas já chegou ao máximo.

No projeto foram escolhidos como valores de controle **500 páginas no máximo e uma profundidade de 4** *urls*.

PageFetcher

Classe que representa as threads de busca do coletor. A classe possui 2 atributos:

- escalonador: Contém o escalonador que é compartilhado por todas as threads.
- robotName: Contém o nome do coletor (nesse projeto, o asyrobot).

As *threads* são responsáveis por a partir de uma *url* buscar a página da mesma, percorrer essa página buscando os links existentes e, para cada link, adicionar (ou não) a url do link à lista do escalonador. Os métodos da classe utilizados para esse fim são:

- coletaLinksHtml: Método responsável por resgatar a página da url, percorrer a mesma. Para cada link encontrado na página, caso não seja do tipo php e contenha o domínio do servidor pertencente, adiciona a mesma à lista de urls, indicando a profundidade da mesma como consecutiva da url utilizada para buscar. A busca da página utiliza a biblioteca <u>HtmlCleaner</u>, a qual possui métodos que buscam e ajustam htmls mal formatados.
- **getRecord**: Método responsável por buscar a configuração do robots.txt do servidor cuja url está sendo buscada.
- run: Método principal da thread, sua duração executa até que o escalonador sinalize que a coleta foi finalizada. Busca a próxima url na fila do escalonador. Inicialmente, verifica quais as regras de busca do servidor (quando ainda não existe no mapeamento do escalonar, busca no servidor e adiciona ao mapeamento), e caso a url respeite todas as regras coleta as urls da página adicionando à fila.

Devido à forma como a coleta ocorre, existe uma execução contínua pois cada url gera mais links que servem para continuar a execução. Dado que o escalonador é compartilhado, quando a execução termina para um *thread*, termina para todas.

Urls Coletadas

http://www.zerozero.pt

http://terra.com.br

http://www.forbes.com https://www.ceroacero.es https://www.ogol.com.br https://www.zerozero.pt https://www.thefinalball.com

https://www.calciozz.it https://www.leballonrond.fr https://www.fussballzz.de https://www.soccerzz.com

https://facebook.com/zerozeropt https://twitter.com/zerozeropt https://instagram.com/zerozeropt https://www.youtube.com/c/zerozero

http://wlbetpt.adsrv.eacdn.com/C.ashx?btag=a 45b 46c &affid=45&siteid=45&adid=46&c=

http://www.sporttv.pt

http://elevensports.com/pt-pt/

https://www.casinoportugal.pt/promocoes?btag=900498&ad=vSmgOLPz9pqO0r3b

https://www.facebook.com/ogolcombr

http://refpa.top/L?tag=d 27572m 97c &site=27572&ad=97

https://www.instagram.com/ogolcombr/

http://www.bet365.com/dl/~offer?affiliate=365_033182 https://www.bet.pt/apostas-desportivas/?btag=a_45b_46c_

http://facebook.com/zerozeropt https://twitter.com/playmaker_PT http://instagram.com/zerozeropt http://www.youtube.com/c/zerozero

https://www.sporttv.pt/

https://www.casinoportugal.pt/desportos/mercados/6073440?btag=900498&ad=vSmgOLPz9pgO0r3b

https://pbs.twimg.com/profile_images/943146940593778690/AqeuRWNk_400x400.jpg

http://t.co/TAtJlqwi0D http://status.twitter.com

http://support.twitter.com/forums/26810/entries/78525

https://dev.twitter.com/web/embedded-tweets

 $http://www.bet365.com/betslip/instantbet/default.aspx?bs=10534311 \sim 2.30 \& Instantbet=1 \& affine the control of the control$

liatecode=365_505393

https://login.sporttv.pt/sso/Login.action?redirectURL=https://www.sporttv.pt/

https://oferta.sporttv.pt/

https://www.facebook.com/SPORTTVPortugal https://www.instagram.com/sporttvportugal/

http://liganos.pt/

http://www.premierleague.com/

https://www.allianz.pt/allianz-cup

https://pt.uefa.com/uefanationsleague/index.html

http://www.legaseriea.it/it

http://www.nba.com/

http://uk.ufc.com/

http://www.motogp.com/pt

http://www.wrc.com/

http://www.nascar.com/en_us/sprint-cup-series.html

http://www.wwe.com/

http://www.getfirefox.com/

http://www.google.com/chrome/

http://www.getie.com/

http://www.opera.com/

http://www.apple.com/safari/

https://www.facebook.com/

https://es-la.facebook.com/

https://fr-fr.facebook.com/

https://it-it.facebook.com/

https://de-de.facebook.com/

https://ar-ar.facebook.com/

https://hi-in.facebook.com/

https://zh-cn.facebook.com/

https://ja-jp.facebook.com/ https://messenger.com/

https://developers.facebook.com/?ref=pf

https://www.getfirefox.com/

https://support.mozilla.org/kb/update-latest-version-firefox-android

https://download.mozilla.org/?product=firefox-stub&os=win64&lang=en-US

https://play.google.com/store/apps/details?id=org.mozilla.firefox&referrer=utm source%

3Dmozilla%26utm_medium%3DReferral%26utm_campaign%3Dmozilla-org

https://itunes.apple.com/us/app/apple-store/id989804926?pt=373246&mt=8

https://accounts.firefox.com/legal/terms

https://blog.mozilla.org/press/

https://careers.mozilla.org/?utm_source=www.mozilla.org&utm_medium=referral&utm_camp aign=footer&utm_content=company

https://developer.mozilla.org/?utm_source=www.mozilla.org&utm_medium=referral&utm_campaign=footer&utm_content=developers

https://bugzilla.mozilla.org/

https://www.instagram.com/mozilla/

https://foundation.mozilla.org

https://facebook.com/ogolcombr

http://twitter.com/zerozeropt

https://instagram.com/ogolcombr

https://www.youtube.com/channel/UCjWcRCWje4HvkD63qqIM8JQ

https://www.casinoportugal.pt/desportos/mercados/6072342?btag=900498&ad=vSmgOLPz9pqO0r3b

https://pbs.twimg.com/profile_images/943146940593778690/AgeuRWNk.jpg

https://t.co/luBrEISZf1

http://support.twitter.com/articles/14226-how-to-find-your-twitter-short-code-or-long-code

https://dev.twitter.com/overview/terms/agreement

http://www.bet365.com/betslip/instantbet/default.aspx?bs=10534312~3.50&Instantbet=1&affiliatecode=365 505393

http://pt.uefa.com/uefaeuropaleague/index.html

https://help.periscope.tv/customer/en/portal/articles/2600293-what-is-periscope-producer

https://www.twitter.com

https://www.periscope.tv/trademarkpolicy https://support.twitter.com/articles/76915

https://help.twitter.com/en/rules-and-policies/twitter-cookies

https://policies.google.com/privacy?hl=en https://www.linkedin.com/legal/privacy-policy https://www.newscred.com/privacypolicy/

https://about.twitter.com

https://blog.twitter.com/official/en us.html

https://careers.twitter.com/en.html https://investor.twitterinc.com/ https://business.twitter.com/

https://www.facebook.com/recover/initiate?lwv=110&ars=royal blue bar

https://ads.twitter.com?ref=en-btc-gobal-footer

https://twittercommunity.com/ https://marketing.twitter.com/

https://takeflight.twitter.com/?utm_source=twitter&utm_medium=organic&utm_campaign=20

190305 global fs Twitter-Footer-Promotion

https://twittercommunity.com/c/dev-platform-feedback/rules-and-policies

https://es-la.facebook.com/login/identify/?ctx=recover https://fr-fr.facebook.com/login/identify/?ctx=recover https://it-it.facebook.com/login/identify/?ctx=recover https://de-de.facebook.com/login/identify/?ctx=recover https://ar-ar.facebook.com/login/identify/?ctx=recover

https://hi-in.facebook.com/login/identify/?ctx=recover https://zh-cn.facebook.com/login/identify/?ctx=recover

https://ja-jp.facebook.com/login/identify/?ctx=recover

https://ads.twitter.com/login?ref=en-btc-gobal-footer&lang=ar

https://support.twitter.com/groups/58-advertising

http://status.twitter.com/

https://www.google.com/?hl=en

http://privacy?hl=en http://terms?hl=en

http://technologies?hl=en

http://faq?hl=en

https://myaccount.google.com/?hl=en

https://www.gstatic.com/policies/privacy/pdf/20190122/f3294e95/google_privacy_policy_en.p df

https://support.google.com/policies?p=privpol_privts&hl=en

https://adssettings.google.com/?utm_source=pp&hl=en

https://aboutme.google.com/?utm_source=pp&hl=en

https://plus.google.com/settings?utm_source=pp&hl=en

https://myactivity.google.com/myactivity?utm_source=pp&hl=en

https://takeout.google.com/?utm_source=pp&hl=en

https://transparencyreport.google.com/user-data/overview?hl=en

https://safebrowsing.google.com/?utm_source=pp&hl=en

https://play.google.com/books/intl/en/privacy.html

https://payments.google.com/legaldocument?family=0.privacynotice&hl=en

https://fiber.google.com/legal/privacy.html

https://fi.google.com/about/tos/

https://kids.youtube.com/privacynotice

https://families.google.com/familylink/privacy/child-policy/

https://contacts.google.com?hl=en

https://privacy.google.com/businesses/affiliates?hl=en

https://support.mozilla.org/kb/update-firefox-latest-version

https://download.mozilla.org/?product=firefox-stub&os=win&lang=en-US

https://itunes.apple.com/us/app/apple-store/id989804926?pt=373246&mt=8&ct=m

ozorg-fxnew page scene1 modal-appstore-button

https://accounts.firefox.com/legal/privacy

https://developer.mozilla.org/docs/Tools/?utm_source=www.mozilla.org&utm_medium=referr

al&utm campaign=footer&utm content=developers

https://facebook.com/ceroacero.es

https://twitter.com/ogolcombr

http://www.youtube.com/subscription center?add user=sporttv

https://www.casinoportugal.pt/desportos/mercados/5394221?btag=900498&ad=vSmgOLPz9pqO0r3b

https://pbs.twimg.com/profile_images/606967013634916352/hzqsFzdl_400x400.png

https://t.co/V3GXgBvLPh

http://www.bet365.com/betslip/instantbet/default.aspx?bs=10534313~3.00&Instantbet=1&affi

liatecode=365 505393

https://dev.twitter.com/overview/terms/policy

https://pt.uefa.com/european-qualifiers/index.html

https://www.twitter.com/privacy/

https://help.twitter.com/

https://about.twitter.com/en us/lets-go-twitter.html

https://blog.twitter.com/small-business

https://support.twitter.com/articles/20172501

https://business.twitter.com/en/advertising.html

https://www.periscope.tv/content

https://www.facebook.com/SPORTTVPortugal/likes?ref=page internal

https://marketing.twitter.com/na/en/insights.html

https://help.periscope.tv/customer/portal/emails/new

https://store.uefa.com/

https://plus.google.com/+UEFAcom https://instagram.com/uefacom/

https://www.linkedin.com/company/uefa

https://uefa.newsmarket.com/

https://pt.uefa.com/european-qualifiers/video/highlights/

http://../uefaeuro-2020/performancezone

http://www.uefafoundation.org http://row.store.uefa.com/

https://itunes.apple.com/pt/app/id1061115611?mt=8

https://play.google.com/store/apps/details?id=com.uefa.euro2016

http://ac.uefa.com/pt/

https://twitter.tlnk.io/serve?action=click&publisher_id=359298&site_id=64980&sub_site=Lets

GoTwitter&site id android=64602&my campaign=LetsGoTwitter

https://mobile.twitter.com/explore

http://www.so-netmedia.jp/privacypolicy/

https://ads.twitter.com/login?ref=en-btc-gobal-footer&lang=fr

https://medium.com/@periscope

https://www.google.com/history/optout?utm_source=pp&hl=en

http://privacy/frameworks?hl=en

http://technologies/product-privacy?hl=en

https://careers.twitter.com/en/teams/periscope.html

https://help.pscp.tv

https://myaccount.google.com/?utm_source=pp&hl=en

https://support.google.com/accounts?p=privpol_location&hl=en

https://myactivity.google.com/?utm source=pp&hl=en

https://www.periscope.tv/i/contact https://tweetdeck.twitter.com/

https://vine.co/privacy

https://www.niche.co/privacy

https://www.snappytv.com/privacy

https://www.mopub.com/legal/privacy/

https://twitpic.com/static/privacy

https://accounts.google.com/signin/usernamerecovery?continue=https%3A%2F%2Fplus.goo

gle.com%2F%2BUEFAcom&osid=1&hl=pt

https://support.apple.com/pt-pt

https://www.apple.com/pt/ios/app-store/

https://helpcenter.uefa.com/hc/en-gb/requests/new

https://locate.apple.com/pt/pt/

https://www.google.com.br/webhp?tab=8w

https://maps.google.com.br/maps?hl=en&tab=8l

https://mail.google.com/mail/?tab=8m

https://drive.google.com/?tab=8o

https://www.uefa.com/

http://uefa.com

http://developer.android.com/index.html

https://payments.google.com/termsOfService?hl=en US

http://./

http://pt.user.uefa.com/Account/SignUp

https://support.mozilla.org/kb/end-support-windows-xp-and-vista

https://download.mozilla.org/?product=firefox-latest-ssl&os=osx&lang=en-US

 $https://accounts.firefox.com/?action=email\& service=sync\& context=fx_desktop_v3$

&entrypoint=accounts-page&utm_content=%2Ffirefox%2Fnew%2F&utm_sou

rce=download-page&utm_campaign=fxa-embedded-form

https://facebook.com/pages/calciozzit/244114205627170

https://twitter.com/ceroacero

https://www.youtube.com/feed/history/search history?utm source=pp&hl=en

https://www.casinoportugal.pt?btag=900498&ad=vSmgOLPz9pqO0r3b

https://pbs.twimg.com/profile_images/606967013634916352/hzqsFzdl.png

https://t.co/veKz6Hx6nt

http://www.bet365.com/betslip/instantbet/default.aspx?bs=10534324~3.10&Instantbet=1&affi

liatecode=365 505393

https://dev.twitter.com/web/embedded-timelines

https://help.twitter.com/using-twitter

https://about.twitter.com/en_us/company.html

https://blog.twitter.com/developer

https://support.twitter.com/forms/platform

https://business.twitter.com/en/targeting.html

https://www.facebook.com/SPORTTVPortugal/likes

https://marketing.twitter.com/na/en/success-stories.html

https://developer.twitter.com/en/docs/basics/getting-started

https://pt.uefa.com/european-qualifiers/season=2020/standings/

https://twittercommunity.com/t/update-on-the-embedded-grid-display-type/119564

https://itunes.apple.com/us/app/id972909677?mt=8

https://ads.twitter.com/login?ref=en-btc-gobal-footer&lang=pt

https://play.google.com/store/apps/details?id=com.twitter.android&hl=en

https://medium.com/@periscope/aea99291d998

https://publish.twitter.com/

https://www.google.com/landing/2step/?utm_source=pp&hl=en

http://privacy/key-terms?hl=en

http://technologies/retention?hl=en

https://myaccount.google.com/security-checkup?utm_source=pp&hl=en

https://support.google.com/accounts?p=privpol_lochistory&hl=en

https://help.pscp.tv/customer/portal/emails/new?t=607538

http://www.w3.org/WAI/PF/aria-practices/

https://investor.twitterinc.com/corporate-governance/board-of-directors/default.aspx

amp;osid=1

http://www.google.com.br/imghp?hl=en&tab=8i

https://www.uefa.com/privacypolicy/index.html

https://support.mozilla.org/kb/firefox-osx

https://download.mozilla.org/?product=firefox-latest-ssl&os=linux64&lang=en-US

https://payments.google.com/legaldocument?family=0.privacynotice&hl=en_US

http://pt.user.uefa.com/Account/LogOn

https://careers.twitter.com/ https://engineering.twitter.com https://twittercommunity.com

https://get.fabric.io

https://twitterflightschool.com https://twitter.com/calciozz

https://www.youtube.com/feed/history?utm_source=pp&hl=en

https://pbs.twimg.com/profile_images/606967279121801216/K9Hn8Rwy_400x400.png

http://www.bet365.com/betslip/instantbet/default.aspx?bs=10534325~3.40&Instantbet=1&affi

liatecode=365_505393 https://t.co/0cRiNM1WPZ https://dev.twitter.com/

https://help.twitter.com/managing-your-account https://about.twitter.com/en_us/values.html

https://blog.twitter.com/advertising

https://business.twitter.com/en/analytics.html

https://www.facebook.com/REPORTV/

https://marketing.twitter.com/na/en/solutions.html

https://support.twitter.com/groups/56-policies-violations/topics/236-twitter-rules-policies/articl

es/18311-the-twitter-rules

https://developer.twitter.com/en/docs/basics/things-every-developer-should-know

https://pt.uefa.com/european-qualifiers/season=2020/matches/index.html

https://github.com/twitter/twurl

https://itunes.apple.com/pt/developer/uefa/id326853637?mt=8

https://trello.com/b/myf7rKwV/twitter-api-platform-roadmap

http://twittercommunity.com https://transparency.twitter.com/

https://ads.twitter.com/login?ref=en-btc-gobal-footer&lang=it

https://analytics.twitter.com/

https://www.google.com/chrome/intl/en/privacy.html

https://play.google.com/store/apps/details?id=tv.periscope.android

https://unionmetrics.com/free-tools/twitter-assistant?utm_source=btc&utm_campaign=analyti

cs&utm_content=box

http://privacy/google-partners?hl=en

http://technologies/cookies?hl=en

https://myaccount.google.com/privacycheckup?utm source=pp&hl=en

https://support.google.com/legal?p=privpol_remove&hl=en

https://investor.twitterinc.com

https://accounts.google.com/SignUp?continue=https%3A%2F%2Fplus.google.com%2F%2B

UEFAcom

https://www.google.com.br/intl/en/about/products?tab=8h

https://support.mozilla.org/kb/install-firefox-linux

https://download.mozilla.org/?product=firefox-latest-ssl&os=linux&lang=en-US

https://twitter.com/signup

http://www.youtube.com/UEFA

https://pbs.twimg.com/profile_images/606967279121801216/K9Hn8Rwy.png

http://www.bet365.com/betslip/instantbet/default.aspx?bs=10534326~2.30&Instantbet=1&affi

liatecode=365_505393 https://t.co/Wd2JGrL4Vn

https://dev.twitter.com/overview/documentation

https://help.twitter.com/safety-and-security https://about.twitter.com/en_us/safety.html

http://blog.twitter.com/

https://business.twitter.com/en/help.html https://www.facebook.com/worldrugby7s/

https://marketing.twitter.com/na/en/collections.html

https://support.twitter.com/articles/79901

https://pt.uefa.com/european-qualifiers/news/newsid=2503900.html?iv=true

https://developer.twitter.com/en/docs/basics/FAQ

https://ads.twitter.com/login?ref=en-btc-gobal-footer&lang=nl

https://itunes.apple.com/pt/genre/id6004 https://analytics.twitter.com/user/*/tweets

https://www.google.com/work/apps/terms/education_privacy.html

http://privacy/archive?hl=en http://technologies/ads?hl=en

https://myaccount.google.com/activitycontrols?utm source=pp&hl=en

https://support.google.com/websearch?p=privpol_locserp&hl=en

https://play.google.com/apps

https://unionmetrics.com/free-tools/twitter-assistant?utm_source=btc&utm_campaign=analytics&utm_content=footer

https://accounts.google.com/TOS?loc=BR&hl=pt&privacy=true

https://support.mozilla.org/kb/refresh-firefox-reset-add-ons-and-settings?utm_source=mozilla.org&utm_medium=referral&utm_campaign=fix-problem-link

https://download.mozilla.org/?product=firefox-latest-ssl&os=win64&lang=en-US

https://twitter.com/sporttvportugal

https://www.youtube.com/?gl=BR&tab=81

https://pbs.twimg.com/profile_images/606967478426738688/orshsJwl_400x400.png

http://www.bet365.com/betslip/instantbet/default.aspx?bs=10534365~1.50&Instantbet=1&affi

liatecode=365_505393 https://t.co/6EdHuRn5tk

https://dev.twitter.com/community

https://help.twitter.com/rules-and-policies

https://about.twitter.com/en_us/company/brand-resources.html

https://blog.twitter.com

https://business.twitter.com/en/help/campaign-editing-and-optimization/public-to-private-conversation.html

https://www.facebook.com/uefacom/

https://marketing.twitter.com/na/en/subscribe.html?subscribe cta=subscribe-footer

https://support.twitter.com/articles/14016

https://pt.uefa.com/european-qualifiers/news/newsid=2565113.html

https://developer.twitter.com/en/docs/basics/apps

https://ads.twitter.com/login?ref=en-btc-gobal-footer&lang=ko

https://www.google.com/intl/en/safetycenter/

https://itunes.apple.com/pt/app/uefa-integrity/id918431353?mt=8

https://analytics.twitter.com/user/*/tweets?filter=top

http://technologies/pattern-recognition?hl=en

https://myaccount.google.com/shared-endorsements?utm source=pp&hl=en

https://support.google.com/accounts?p=privpol_endorse&hl=en

https://play.google.com/store/apps/category/GAME

https://accounts.google.com/TOS?loc=BR&hl=pt

https://support.mozilla.org/products/?utm_source=mozilla.org&utm_medium=referral&utm_campaign=need-help-link

https://download.mozilla.org/?product=firefox-latest-ssl&os=win&lang=en-US

https://twitter.com/mozilla

https://pbs.twimg.com/profile_images/606967478426738688/orshsJwl.png

http://www.bet365.com/betslip/instantbet/default.aspx?bs=10534366~4.20&Instantbet=1&affi

liatecode=365_505393 https://t.co/VEYHnhwpL7

https://dev.twitter.com/overview/terms/agreement-and-policy

https://help.twitter.com/contact-us

https://about.twitter.com/company/display-requirements

https://blog.twitter.com/developer/en_us.html

https://www.facebook.com/LaLiga/

https://business.twitter.com/en/blog/podcast.html

https://marketing.twitter.com/emea/en_gb/insights.html

https://support.twitter.com/articles/20169453

http://pt.uefa.com/european-qualifiers/video/highlights/index.html

https://developer.twitter.com/en/docs/basics/developer-portal

https://ads.twitter.com/login?ref=en-btc-gobal-footer&lang=de

https://www.google.com/history/optout?hl=en

https://itunes.apple.com/pt/app/champions-league/id1024506667?mt=8

http://technologies/partner-sites?hl=en

https://myaccount.google.com/dashboard?utm_source=pp&hl=en

https://support.google.com/googlehome?p=privpol homedata&hl=en

https://play.google.com/store/apps/category/FAMILY

https://accounts.google.com/ServiceLogin?passive=1209600&continue=https://play.google.com/store/apps/details?id%3Dcom.uefa.euro2016&followup=https://play.google.com/store/apps/details?id%3Dcom.uefa.euro2016

https://support.mozilla.org/products/mobile/?utm_source=mozilla.org&utm_medium=referral&utm_campaign=need-help-link

https://twitter.com/login

https://pbs.twimg.com/profile_images/979315330836254720/kVmaZe34_400x400.jpg

http://www.bet365.com/betslip/instantbet/default.aspx?bs=10534367~6.50&Instantbet=1&affi

liatecode=365_505393

https://t.co/gq5g9iHJdE

https://dev.twitter.com/oauth

https://help.twitter.com/rules-and-policies/twitter-cookies

https://about.twitter.com/content/dam/about-twitter/company/executive-leadership/Parag-Aga

rwal_HiRes.jpg

https://blog.twitter.com/en_us/topics/company/2019/health-update.html

https://www.facebook.com/wimbledon/

https://business.twitter.com/en/blog/having-personal-Twitter-account-steps-up-your-ad-game

.html

https://marketing.twitter.com/emea/fr/insights.html

https://developers.google.com/analytics/devguides/collection/analyticsjs/cookie-usage

https://support.twitter.com/articles/18311

http://pt.uefa.com

http://optout.aboutads.info/

http://optout.networkadvertising.org/

https://tools.google.com/dlpage/gaoptout

https://ads.twitter.com/login?ref=en-btc-gobal-footer&lang=ru

https://www.google.com/about/datacenters/inside/locations?hl=en

https://developer.twitter.com/en/docs/basics/authentication

https://itunes.apple.com/pt/app/uefa-europa-league/id1048183714?mt=8

http://technologies/types?hl=en

https://myaccount.google.com/privacy?utm source=pp&hl=en

https://support.google.com/a?p=privpol_admin&hl=en

https://play.google.com/store/apps/editors_choice

https://support.mozilla.org/products/ios/?utm_source=mozilla.org&utm_medium=referral

&utm campaign=need-help-link

https://twitter.com/logo

https://pbs.twimg.com/profile_images/979315330836254720/kVmaZe34.jpg

http://www.bet365.com/betslip/instantbet/default.aspx?bs=10534337~1.72&Instantbet=1&affi

liatecode=365_505393 https://t.co/BmyDfCOSRG

http://dev.twitter.com/cards

https://help.twitter.com/en/using-twitter

https://www.facebook.com/WorldAthleticsClub/

https://about.twitter.com/content/dam/about-twitter/company/executive-leadership/Leslie-Berl

and HiRes.jpg

https://blog.twitter.com/en_us/authors.dhicks.html

https://business.twitter.com/en/blog/ webinar-recap -7-keys-to-twitter-success.html

https://marketing.twitter.com/emea/ar/insights.html

https://support.twitter.com/articles/114233

https://pt.uefa.com/

https://ads.twitter.com/login?ref=en-btc-gobal-footer&lang=th

https://www.google.com/

https://developer.twitter.com/en/docs/basics/rate-limits

https://itunes.apple.com/pt/app/central-de-jogos-champions-le/id1107602457?mt=8

http://technologies/anonymization?hl=en

https://myaccount.google.com/deleteservices?utm_source=pp&hl=en

https://support.google.com/accounts?p=privpol_whyad&hl=en

https://play.google.com/store/movies

https://support.mozilla.org/?utm_source=www.mozilla.org&utm_medium=referral&utm_camp aign=footer&utm_content=help

http://twitter.com/twitterdev

http://www.bet365.com/betslip/instantbet/default.aspx?bs=10534338~3.50&Instantbet=1&affi

liatecode=365_505393

https://t.co/Pxncyzlcjz

https://dev.twitter.com/discussions/analytics

https://www.facebook.com/FIVB.InternationalVolleyballFederation/

https://help.twitter.com/en/managing-your-account

https://about.twitter.com/content/dam/about-twitter/company/executive-leadership/Jack-Dors

ey_HiRes.jpg

https://blog.twitter.com/en_us/authors.gasca.html

https://business.twitter.com/en/blog/build-B2B-thought-leadership.html

https://marketing.twitter.com/emea/de/insights.html

https://support.twitter.com/articles/20170514

http://pt.uefa.com/privacypolicy.html

https://ads.twitter.com/login?ref=en-btc-gobal-footer&lang=id

https://www.google.com/about/

https://itunes.apple.com/pt/app/uefa-go/id1114696441?mt=8

https://developer.twitter.com/en/docs/basics/rate-limiting

https://myaccount.google.com/deleteaccount?utm_source=pp&hl=en

https://support.google.com/accounts?p=privpol_phone&hl=en

https://play.google.com/movies

https://twitter.com/privacy

https://cdn.cms-twdigitalassets.com/content/dam/legal-twitter/site-assets/privacy-page-gdpr/

pdfs/PP_Q22018_April_EN.pdf https://pscp.tv/account/settings

https://help.pscp.tv/customer/portal/articles/2017789

https://optout.aboutads.info/

https://www.pscp.tv/account/settings

https://twitter.ethicspointvp.com/custom/twitter/forms/data/form_data.asp

https://www.dataprotection.ie/docs/Contact-us/b/11.htm

https://www.privacyshield.gov/participant?id&

https://www.jamsadr.com/eu-us-privacy-shield

http://www.bet365.com/betslip/instantbet/default.aspx?bs=10534339~5.00&Instantbet=1&affi

liatecode=365_505393 https://t.co/Su9pdhuKai

https://dev.twitter.com/twitterkit/overview

https://www.facebook.com/tonline.fussball/

https://help.twitter.com/en/safety-and-security

https://about.twitter.com/content/dam/about-twitter/company/executive-leadership/Bruce-Fal

ck HiRes.png

https://blog.twitter.com/en_us/topics/product/2018/more-clarity-on-reported-tweets-and-enfor

cement.html

https://business.twitter.com/en/blog/how-video-is-reshaping-digital-advertising.html

https://marketing.twitter.com/emea/es_es/insights.html

https://github.com/twitter/twitter-kit-android

https://support.twitter.com/articles/20175256

http://pt.uefa.com/community/index.html

https://ads.twitter.com/login?ref=en-btc-gobal-footer&lang=sv

https://www.google.com/intl/pt/about

https://myaccount.google.com/preferences?utm_source=pp&hl=en

https://itunes.apple.com/pt/app/uefa-womens-football/id1220891855?mt=8

https://support.google.com/accounts?p=privpol_agereq&hl=en

https://play.google.com/store/movies/category/FAMILY

https://developer.twitter.com/en/docs/basics/response-codes

https://twitter.com/tos

https://help.pscp.tv/customer/portal/articles/2017791-how-do-i-share-my-location-

https://www.privacyshield.gov/article?id&

http://www.bet365.com/betslip/instantbet/default.aspx?bs=10534352~3.00&Instantbet=1&affi

liatecode=365_505393 https://t.co/Hzc9t4YMHf

https://dev.twitter.com/web/overview https://www.facebook.com/PGATour/

https://help.twitter.com/en/rules-and-policies

https://about.twitter.com/content/dam/about-twitter/company/executive-leadership/Vijaya-Ga

dde_HiRes.jpg

https://blog.twitter.com/en_us/authors.toizer.html

https://business.twitter.com/en/blog/Character-Count-podcast-Monterey-Bay-Aquarium.html

https://marketing.twitter.com/latam/pt/insights.html

https://support.twitter.com/forms/dmca

http://pt.uefa.com/store/index.html

https://github.com/twitter/twitter-kit-ios

https://ads.twitter.com/login?ref=en-btc-gobal-footer&lang=tr

https://publish.twitter.com

https://twittercommunity.com/c/publisher/websites

http://www.google.com/support/accounts?hl=pt

https://myaccount.google.com/inactive?utm_source=pp&hl=en

https://support.google.com/chrome?p=privpol_chrsync&hl=en

https://itunes.apple.com/pt/app/forza-football/id500138120?mt=8

https://play.google.com/store/movies/stream/promotion collections movie studios

https://developer.twitter.com/en/docs/basics/cursoring

https://twitter.com

https://help.pscp.tv/customer/portal/articles/2932270

https://www.privacyshield.gov/welcome

http://www.bet365.com/betslip/instantbet/default.aspx?bs=10534353~3.30&Instantbet=1&affi

liatecode=365_505393 https://t.co/IdOglXMuwq

https://dev.twitter.com/web/overview/privacy

https://www.facebook.com/NHL/

https://help.twitter.com/en/twitter-guide

https://about.twitter.com/content/dam/about-twitter/company/executive-leadership/Ned-Segal _HiRes.jpg

https://blog.twitter.com/en_us/topics/company/2018/Creating-new-policies-together.html https://business.twitter.com/en/help/overview/ads-pricing.html?utm_source=banner&utm_me dium=pricing

https://marketing.twitter.com/latam/es/insights.html

https://ads.twitter.com/login?ref=en-btc-gobal-footer&lang=da

https://support.twitter.com/articles/18367