

Coleta de Dados do Instagram sobre Ocorrências de Caravelas-Portuguesas na Costa Brasileira

Leonardo da S. Camargo¹, Heloisa Rocha¹, Lorena S. do Nascimento², Carmem Hara¹

¹Departamento de Informática, Curitiba-PR

²Centro de Estudos do Mar, Pontal do Paraná-PR
Universidade Federal do Paraná

{camargo.s.leonardo, heloisarocha, lorena.sn, carmemhara}@ufpr.br

Abstract. *Social media generate data in large volumes, which are freely and easily accessible, renewable, because they are generated continuously and in real time, and are long-lasting. The objective of this work is to collect posts from Instagram in order to map the spatio-temporal distribution of sightings of the Portuguese man of war (cnidary physalia physalis) on the Brazilian coast. The amount of collected posts indicates that Instagram is a potential data source for obtaining this type of data. Future work involves determining the veracity and scope of the collected information and developing an automated process for periodically extracting new posts.*

Resumo. *Redes sociais geram dados em grande volume, que são de livre e fácil acesso, renováveis por serem gerados continuamente e em tempo real, e de longa duração. O objetivo deste trabalho é coletar postagens do Instagram para obter dados sobre a distribuição espaço-temporal de avistamentos da caravela-portuguesa (cnidário physalia physalis) na costa brasileira. Foram obtidas postagens do Instagram com informações que apontam o Instagram como uma possível fonte deste tipo de dado, dada a quantidade de postagens coletadas. Trabalhos futuros envolvem determinar a veracidade e abrangência das informações e desenvolvimento de um processo automatizado para periodicamente extrair novas postagens.*

1. Introdução

Um número sem precedentes de dados estão disponíveis nas redes sociais e podem ser utilizados para os estudos ecológicos (Jarić et al., 2020). São diversos os trabalhos na área de ecologia que utilizam dados coletados de redes sociais, tais como: monitoramento de animais marinhos (Paolo et al., 2021), peixes de rio (Al Mabruk et al., 2021), estudos que ajudam na conservação das espécies (Rizgalla, 2021) e também delimitação geográfica das espécies (Kovačić et al., 2023).

Um ponto que vale destacar sobre os dados fornecidos pelas redes sociais é que são de livre e fácil acesso. Além disso, eles são renováveis, por serem gerados continuamente e em tempo real, são de longa duração e são gerados em grande volume (Di Minin et al., 2015; Toivonen et al., 2019). Existem estudos que, ao comparar os dados obtidos em redes sociais com outras fontes, como dados online e registros da literatura, concluem que as informações obtidas sobre a distribuição espaço-temporal possuem, em sua maioria, níveis satisfatórios de consistência (Barve, 2014; Ghermandi and Sinclair, 2019).

Tendo em vista os resultados obtidos pelos autores citados, este trabalho tem como objetivo coletar postagens da rede social Instagram, a fim de criar uma base de dados contendo informações sobre avistamentos da caravela-portuguesa (i.e. cnidário *Physalia physalis*) na costa brasileira. O início do trabalho teve como base a proposta da tese de doutorado da oceanógrafa do Nascimento et al. (2022), intitulada "Redes sociais como uma fonte de dados alternativa para monitorar águas-vivas no Brasil", que propõe a utilização de dados extraídos manualmente do Instagram para identificação de ocorrências de caravela-portuguesa no litoral brasileiro.

O Instagram permite a associação de postagens a *hashtags*, tais como: #caravela-portuguesa e #aguaviva. As *hashtags* podem ser com palavras existentes ou não, e sempre é possível criar uma nova *hashtag*. *Hashtags* definem categorias de postagens, sendo possível fazer uma busca de postagens por categoria. No entanto, é importante observar que as postagens coletadas não necessariamente são referentes ao objeto de interesse. Por exemplo a *hashtag* #caravelaportuguesa pode ser associada ao animal de interesse do trabalho ou a uma embarcação inventada pelos portugueses. Mesmo quando a postagem refere-se ao animal de interesse, ela nem sempre possui os dados necessários para o mapeamento espaço-temporal das ocorrências na costa litorânea brasileira. Dessa forma, o resultado final do processo de coleta automatizado descrito neste artigo é a geração de uma planilha, na qual cada postagem coletada pode manualmente ser rotulada como positivo ou negativo para o avistamento da espécie.

Este trabalho está dentro do contexto de um projeto mais abrangente, que tem por objetivo desenvolver técnicas para automatizar o processo de coleta contínua, análise (Rocha and Hara, 2022) e integração com outras fontes de dados (Reips and Hara, 2022) a respeito da caravela-portuguesa. Dessa forma, as planilhas com dados rotulados podem ser utilizados para o treinamento de modelos de aprendizagem de máquina para a etapa de análise, a fim de automaticamente identificar postagens com ocorrências verdadeiras de caravela-portuguesa.

O restante do artigo está organizado da seguinte forma: a Seção 2 descreve trabalhos relacionados. A Seção 3 descreve o processo e ferramentas utilizadas na coleta e geração de planilhas. O artigo conclui na Seção 4 com algumas considerações finais.

2. Trabalhos Relacionados

No meio acadêmico, diversos trabalhos utilizam o Twitter como fonte de dados para identificar observações da vida selvagem em redes sociais, tal como o de Edwards et al. (2022). Esta rede social até recentemente possuía uma forma fácil e gratuita para extração de dados através de uma API. Entretanto, há mudanças neste cenário, com novas políticas que dificultam o acesso aos dados desta rede social. Já a utilização do Instagram como ferramenta de pesquisa não é muito comum no meio acadêmico, embora alguns trabalhos possam ser encontrados. Por exemplo, em Hausmann et al. (2018), os autores utilizam o Instagram e o Flickr para estudar preferências de turistas em experiências naturais de áreas protegidas e ressalta a abundância de dados das plataformas. Existem também casos de estudos de ciência cidadã e redes sociais para delimitação geográfica de espécies (Kovačić et al., 2023). Um fator a ser destacado no Instagram é que ele é uma rede social em que postagens podem disponibilizar informações geoespaciais, ao contrário do Twitter. No entanto, a captura e gerenciamento de dados do Instagram é desafiador de-

vido a mudanças contínuas de políticas de privacidade da plataforma (McCrow-Young, 2021). É possível notar um histórico de limitações cada vez maiores no acesso aos dados (Constine, 2015). Em 2018, houve mudanças na *API*, com uma redução significativa no limite de postagens extraídas utilizando a *API* (Constine, 2018). Em 2020, o Instagram desabilitou permanentemente acesso à aplicação *Legacy* da *API*, resultando em dois anos de paulatina restrição de obtenção de dados por aplicações de terceiros (McCrow-Young, 2021). No final de 2022, a Meta - empresa responsável pelo Instagram - lançou uma nova *API*, que está ainda mais restritiva e que foca em disponibilizar serviços de gerenciamento de contas para empresas e criadores de conteúdo, como análise de dados da própria conta e respostas automáticas a comentários e mensagens.

Atualmente, dentre as possíveis formas de obtenção dos dados estão: a utilização de ferramentas disponibilizadas pelo próprio Instagram (i.e., *CrowdTangle* e *Instagram Graph API*), utilização de *APIs* de terceiros e *Web scraping*, que é um método de extrair caracteres textuais e outras informações de *Web sites*, apesar de não ser exclusivo a apenas esta plataforma (Milev et al., 2017). É possível observar que, apesar das restrições citadas por McCrow-Young (2021), existem autores que utilizam as ferramentas disponibilizadas oficialmente para obtenção dos dados (Hausmann et al., 2018; Gunter and Önder, 2021) e outros que optam pelo desenvolvimento de *Web scraping* próprio (Himawan et al., 2020).

Para o monitoramento de caravelas-portuguesas, as informações mais relevantes são: a localização, mídias (i.e., imagens e vídeos), a legenda e a data em que foi detectada a presença da caravela-portuguesa. A *API CrowdTangle* não permite a obtenção de dados de geolocalização da postagem para que pudesse ser considerada como uma possível candidata para a coleta dos dados. Com as crescentes limitações da *API* do Instagram no período recente (McCrow-Young, 2021; Constine, 2015, 2018), houve dificuldade na adaptação da *API* por conta dessas restrições. A escolha então foi utilizar um *Web scraper* de terceiros, chamada de *Instaloader* (2023). Ele é um *web scraper* implementado em Python 3 que coleta todos os dados relevantes para este trabalho.

3. Coleta de Dados

O processo adotado para a coleta de dados do Instagram está ilustrado na Figura 1. A ferramenta utilizada para a obtenção dos dados do Instagram foi o *Instaloader* (2023). Ele é um software implementado em Python 3 que por meio de *Web scraping* extrai dados como: data de publicação, latitude, longitude, legenda e mídias das postagens. É possível executá-lo por meio de um script em Python ou em linha de comando, utilizando como parâmetro as *hashtags* associadas às postagens. O *Instaloader* (2023) obtém as postagens sempre em ordem da data de postagem mais recente à mais antiga. Vale destacar que o Instagram possui um limite imprevisível para *Web scraping*, sendo que a carga máxima de download muda para cada tentativa de coleta dos dados.

Os dados coletados são organizados pela ferramenta em subdiretórios, sendo criado um para cada *hashtag*. Dentro de cada subdiretório ficam os arquivos referentes às postagens coletadas pela ferramenta. São criados os seguintes arquivos: um arquivo em formato textual (.txt) que contém a legenda da postagem; um arquivo em formato JPEG (.jpg) para cada imagem da postagem; um arquivo em formato MP4 (.mp4) para cada vídeo da postagem e um arquivo em formato JSON (.json) que contém os metadados da postagem (e.g., código da postagem, legenda, data da postagem, latitude, longi-

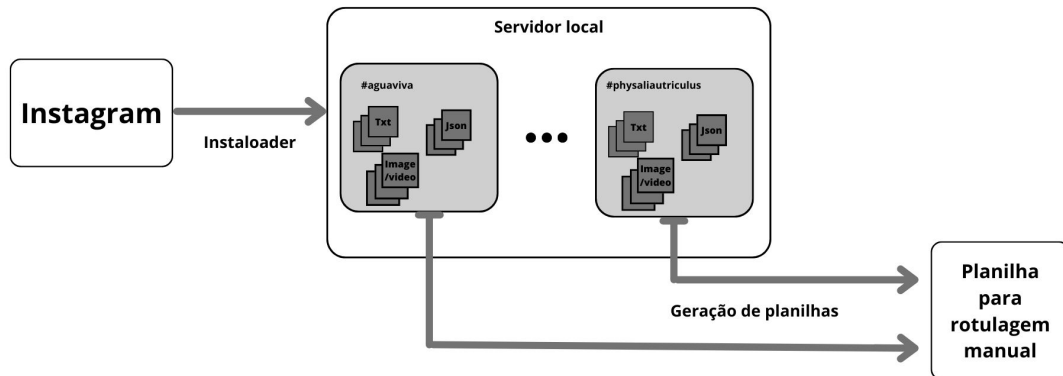


Figura 1. Ilustração do fluxo da coleta de dados

tude). Os diferentes formatos de arquivo de uma postagem se relacionam por terem o mesmo nome. Por exemplo, uma postagem no diretório da #caravelaportuguesa é nomeada pelo *timestamp* da postagem. Então o nome do arquivo JSON é *2012-08-11_20-19-56_UTC.json*, a imagem é *2012-08-11_20-19-56_UTC.jpg* e o texto *2012-08-11_20-19-56_UTC.txt*. Caso exista mais de uma imagem para uma mesma postagem, os nomes das imagens são numerados: por exemplo, *2012-08-11_20-19-56_UTC_1.jpg* e *2012-08-11_20-19-56_UTC_2.jpg*. A Figura 2 ilustra como os dados estão armazenados.

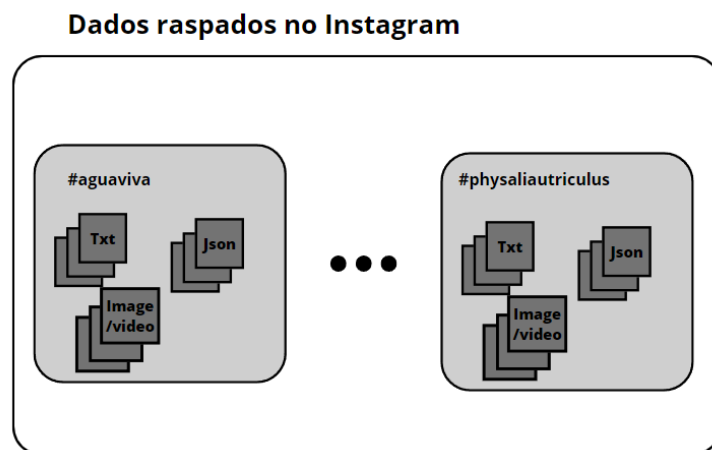


Figura 2. Ilustração do armazenamento de dados coletados

Foi realizado pela oceanógrafa que participa do projeto, um levantamento das *hashtags* que poderiam estar associadas à caravela-portuguesa. Foram considerados: nome científico da espécie, classe, filo, nomes populares e variações em outras línguas. A Tabela 1 mostra o resultado deste levantamento.

Tabela 1. Tabela de *hashtags* candidatas para coleta

águaviva	aguaviva	águasvivas
aguasvivas	cnidario	cnidário
cnidários	bluebottle	bluebottles
jellyfish	jellyfishes	jelly
hydrozoa	hydrozoan	hydrozoans
caravelaportuguesa	carabelaportuguesa	carabelasportuguesas
caravelasportuguesas	bluebottlejellyfish	notajellyfish
medusa	medusas	carabela
manowarjellyfish	caravela	fragataportuguesa
sinophore	caravelas	fragatasportuguesas
sinophores	portuguesemanofwar	garrafaazul
mãedágua	portuguesemanowar	physaliaphysalis
maedagua	portuguesemanofwar	physalia
cnidaria	portuguesemanowar	physaliidae
cnidarios	manofwarjellyfish	physaliautriculus

É importante observar que nem todas as postagens obtidas através destas *hashtags* são realmente do cnidário caravela-portuguesa. Por exemplo, elas podem ser sobre embarcações (navio português), tatuagens, ilustrações, etc. Dessa forma, após uma interação inicial com as *hashtags* acima, a escolha das *hashtags* foi filtrada para aquelas que possuíram um melhor resultado em relação ao objetivo de realizar um mapeamento da distribuição espaço-temporal das ocorrências de caravelas-portuguesas na costa brasileira. A Tabela 2 detalha quais foram as *hashtags* escolhidas, qual o potencial para coleta (consulta realizada em 21/03/2023), quantas foram coletadas e o período coletado. A *#aguaviva*, por exemplo, não possui uma porcentagem de registros de caravela-portuguesa alta, visto que quem classifica uma caravela-portuguesa com esta *hashtag* normalmente o faz por confundir o animal com uma água-viva. Entretanto, o volume de dados em potencial para coleta é muito grande e se mostrou promissor para esta pesquisa. As variações de *#cnidário* (cnidario, cnidarios, cnidários e cnidário) também são abrangentes, visto que possuem muitos animais desse filo que não são caravelas-portuguesas, mas também possuem um grande número de dados que podem ser coletados. *#caravelaportuguesa* e *#physaliaphysalis* tem um potencial alto para obtenção de ocorrências verdadeiras, entretanto o volume de dados disponíveis para coleta é menor.

Tabela 2. Tabela de *hashtags* escolhidas

<i>Hashtag</i>	Nº Potencial para Coleta	Nº de Postagens Coletadas	Período Coletado
<i>#aguaviva</i>	164.124	1.680	07/2021 e 12/2021
<i>#caravelaportuguesa</i>	3.598	1.786	04/2012 e 05/2021
<i>#cnidario</i>	915	528	06/2012 e 03/2021
<i>#cnidário</i>	66	25	04/2013 e 02/2021
<i>#cnidarios</i>	3.462	1.766	12/2014 e 05/2021
<i>#cnidários</i>	482	230	05/2013 e 02/2021
<i>#physaliaphysalis</i>	2.755	189	01/2018 e 12/2021
Total	175.402	6.204	

A coleta dos dados começou em 2021 e foi até datas mais antigas que não puderam ser padronizadas devido ao comportamento limitante e imprevisível do Instagram. Após a coleta dos dados, foram geradas planilhas para rotulagem manual a partir dos dados obtidos pelas seguintes *hashtags*: #caravelaportuguesa, #cnidario e suas variações, #aguaviva e #physaliaphysalis. A rotina para geração das planilhas foi escrita em Python3. Os atributos das planilhas mudaram algumas vezes ao longo do desenvolvimento do trabalho para que atendessem aos requisitos dos demais trabalhos associados ao projeto. Os atributos da planilha que puderam ser obtidos diretamente dos dados extraídos pela ferramenta Instaloader (2023) foram: data da postagem, localização geográfica, cidade, nome da localização, identificador único do link, texto da legenda, imagens e vídeos. Os únicos dados que podem sempre ser obtidos das postagens são o identificador do link e a data da postagem. Ou seja, é possível que a postagem venha sem imagem, texto ou localização.

Os dados coletados também foram enriquecidos com a identificação do idioma da postagem por meio da utilização da biblioteca Langdetect (2023). Para as postagens que tinham geolocalização (ou seja, latitude e longitude) a *API BigDataCloud* (2023) foi usada para obter mais dados de localização, como: cidade, nome do local, estado e país.

A partir das planilhas geradas, foi realizada a rotulagem manual das postagens por uma oceanógrafa. Essa rotulagem teve como objetivo principal determinar se as postagens são de fato ocorrências legítimas de caravela-portuguesa na costa brasileira e, em caso negativo, apontar o motivo: nenhuma imagem corresponde a uma caravela-portuguesa (mídia), localização não corresponde a costa brasileira (localização) ou os ambos (mídia e localização).

3.1. Discussão

A escolha da abordagem para a coleta de postagens do Instagram não foi uma tarefa simples. Embora a decisão final tenha sido pelo *Web scraper* Instaloader, há a possibilidade da ferramenta ser descontinuada e deixar de funcionar devido a alterações na página Web do Instagram. Em pelo menos uma ocasião durante a extração das postagens, a ferramenta deixou de funcionar devido a tais alterações. Além disso, a interface Web tem limitações de quantidade de postagens e permite apenas a captura de dados em ordem cronológica reversa, da mais recente para a mais antiga. A impossibilidade de obtenção de postagens a partir de uma data arbitrária implica em um consumo de carga de requisições desnecessárias para a obtenção de postagens mais antigas. Isso justifica o fato de algumas *hashtags* terem sido obtidas de forma parcial. Outra dificuldade levantada é que quando há uma grande requisição de dados, a conta criada no Instagram para *Web scraping* pode ser barrada, necessitando solicitações ao Instagram para poder utilizar a conta novamente, sendo o número de requisições máxima não informado pela plataforma.

Dados os problemas apresentados, foram realizadas tentativas de utilização da *API* do Instagram para a extração de dados. Essa *API* foi relançada no final de 2022, mas de forma significativamente mais restrita. Devido a constantes mudanças de políticas de privacidade e termos de uso, além da impossibilidade de obtenção de informações geo-espaciais, a utilização da *API* oficial foi descartada.

Com relação à geração da planilha para rotulação, cabe observar que a sua estrutura mudou diversas vezes até ser encontrada uma versão que pudesse atender os requisitos de todos os trabalhos que utilizarão estes dados. Por exemplo, a rotulagem que a espe-

cialista em oceanografia utilizava para considerar uma ocorrência de caravela-portuguesa no Brasil era diferente daquela que estava sendo desenvolvida para o treinamento do modelo de aprendizagem de máquina. Chegar a um consenso para o problema envolveu discussões para um melhor entendimento do problema e diversas versões de planilhas de rotulagem.

4. Conclusão

Este artigo apresentou o processo de coleta de dados da rede social *Instagram* para o monitoramento de avistamentos da caravela-portuguesa no litoral brasileiro.

Foram descritos os dados coletados, bem como as ferramentas utilizadas para a realização da coleta. Uma das grandes dificuldades do trabalho foi na escolha da abordagem para realizar a coleta, dadas as limitações impostas pela *API* do Instagram, bem como pelas dificuldades inerentes ao processo de *Web scraping*. Para não continuar dependentes do *Instaloader*, atualmente está em desenvolvimento uma ferramenta própria para a realização de *Web scraping* de dados. Além da dependência do *Instaloader*, com a realização de um *web scraper próprio*, há a possibilidade de obter os dados aos poucos e de forma contínua, sendo assim contornada o problema da limitação dos dados pelo Instagram. Outro trabalho em andamento é a implementação de um sistema para fazer a coleta contínua de dados para alimentar o processo de geração automatizada de uma base de avistamentos da caravela-portuguesa.

Referências

- S. Al Mabruk, A. Abdulghani, O. Nour, M. Adel, F. Crocetta, N. Doumpas, P. Kleitou, and F. Tiralongo. The role of social media in compensating for the lack of field studies: five new fish species for the mediterranean egypt. *Journal of Fish Biology*, 99, 03 2021. doi: 10.1111/jfb.14721.
- V. Barve. Discovering and developing primary biodiversity data from social networking sites: A novel approach. *Ecological Informatics*, 24:194–199, 2014. ISSN 1574-9541. doi: <https://doi.org/10.1016/j.ecoinf.2014.08.008>. URL <https://www.sciencedirect.com/science/article/pii/S1574954114001186>.
- BigDataCloud. BigDataCloud. <https://www.bigdatacloud.com/>, 2023. Acessado em 03/03/2023.
- J. Constine. Instagram kills off feed reading apps. *Tech Crunch*, 2015. URL <https://techcrunch.com/2015/11/17/just-instagram/>.
- J. Constine. Instagram suddenly chokes off developers as facebook chases privacy. *Tech Crunch*, 2018. URL <https://techcrunch.com/2018/04/02/instagram-api-limit/>.
- E. Di Minin, H. Tenkanen, and T. Toivonen. Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science*, 3, 09 2015. doi: 10.3389/fenvs.2015.00063.
- L. S. do Nascimento, C. S. Hara, M. N. Junior, and M. Noernberg. Instagram como fonte de dados alternativa no monitoramento da caravelaportuguesa (*physalia physalis*, cnidaria). In *Livro de Memórias do IV SUSTENTARE e VII WIPIS: Workshop internacional de Sustentabilidade, Indicadores e Gestão de Recursos Hídricos*. Even3, 2022. doi: 10.29327/sustentare_wipis_2022.584935.

- T. Edwards, C. B. Jones, and P. Corcoran. Identifying wildlife observations on twitter. *Ecological Informatics*, 67:101500, 2022. ISSN 1574-9541. doi: <https://doi.org/10.1016/j.ecoinf.2021.101500>. URL <https://www.sciencedirect.com/science/article/pii/S1574954121002910>.
- A. Ghermandi and M. Sinclair. Passive crowdsourcing of social media in environmental research: A systematic map. *Global Environmental Change*, 55:36–47, 2019. ISSN 0959-3780. doi: <https://doi.org/10.1016/j.gloenvcha.2019.02.003>. URL <https://www.sciencedirect.com/science/article/pii/S0959378018309920>.
- U. Gunter and I. Önder. An exploratory analysis of geotagged photos from instagram for residents of and visitors to vienna. *Journal of Hospitality & Tourism Research*, 45(2): 373–398, 2021.
- A. Hausmann, T. Toivonen, R. Slotow, H. Tenkanen, A. Moilanen, V. Heikinheimo, and E. Di Minin. Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. *Conservation Letters*, 11(1): e12343, 2018. doi: <https://doi.org/10.1111/conl.12343>. URL <https://conbio.onlinelibrary.wiley.com/doi/abs/10.1111/conl.12343>.
- A. Himawan, A. Priadana, and A. Murdiyanto. Implementation of web scraping to build a web-based instagram account data downloader application. *IJID (International Journal on Informatics for Development)*, 9(2):59–65, Dec. 2020. doi: 10.14421/ijid.2020.09201. URL <https://ejournal.uin-suka.ac.id/saintek/ijid/article/view/09201>.
- Instaloader. Instaloader. <https://github.com/instaloader/instaloader>, 2023. Acessado em 03/03/2023.
- I. Jarić, R. A. Correia, B. W. Brook, J. C. Buettel, F. Courchamp, E. Di Minin, J. A. Firth, K. J. Gaston, P. Jepson, G. Kalinkat, R. Ladle, A. Soriano-Redondo, A. T. Souza, and U. Roll. iecology: Harnessing large online resources to generate ecological insights. *Trends in Ecology Evolution*, 35(7):630–639, 2020. ISSN 0169-5347. doi: <https://doi.org/10.1016/j.tree.2020.03.003>. URL <https://www.sciencedirect.com/science/article/pii/S016953472030077X>.
- M. Kovačić, J. P. Renoult, R. Pillon, M. Bilecenoglu, F. Tiralongo, S. V. Bogorodsky, S. Engin, O. Kovtun, P. Louisy, R. A. Patzner, S. B.-S. Rothman, A. Soldo, and M. B. Yokes. The delimitation of geographic distributions of gobius bucchichi and gobius incognitus (teleostei: Gobiidae). *Journal of Marine Science and Engineering*, 11(3), 2023. ISSN 2077-1312. doi: 10.3390/jmse11030516. URL <https://www.mdpi.com/2077-1312/11/3/516>.
- Langdetect. Langdetect. <https://github.com/Mimino666/langdetect>, 2023. Acessado em 03/03/2023.
- A. McCrow-Young. Approaching instagram data: reflections on accessing, archiving and anonymising visual social media. *Communication Research and Practice*, 7(1):21–34, 2021. doi: 10.1080/22041451.2020.1847820. URL <https://doi.org/10.1080/22041451.2020.1847820>.
- P. Milev et al. Conceptual approach for development of web scraping application for tracking information. *Economic Alternatives*, 3:475–485, 2017.
- F. Paolo, F. P. J. Albarico, P. Jee, B. Albarico, C. Peña, M. Salva, and C.-D. Dong. Using social media platforms to study the ecology and exploitation of mud lobsters in the philippines. *Philippine Journal of Science*, 150, 11 2021. doi: 10.56899/150.6B.20.

- L. Reips and C. Hara. Integração e rotulação automatizada de dados sobre o cnidário *physalia physalis*, usando a geolocalização como referência. In *Anais Estendidos do XXXVII Simpósio Brasileiro de Bancos de Dados*, pages 105–111, Porto Alegre, RS, Brasil, 2022. SBC. doi: 10.5753/sbbd_estendido.2022.21851. URL https://sol.sbc.org.br/index.php/sbbd_estendido/article/view/21851.
- J. Rizgalla. Can social media platforms play a role in sea turtle conservation efforts in libya in times of war and political/economic instability? 27:1–23, 05 2021.
- H. Rocha and C. Hara. Identificação de ocorrências do cnidário *physalia physalis* em dados extraídos de mídias sociais. In *Anais Estendidos do XXXVII Simpósio Brasileiro de Bancos de Dados*, pages 119–125, Porto Alegre, RS, Brasil, 2022. SBC. doi: 10.5753/sbbd_estendido.2022.21853. URL https://sol.sbc.org.br/index.php/sbbd_estendido/article/view/21853.
- T. Toivonen, V. Heikinheimo, C. Fink, A. Hausmann, T. Hiippala, O. Järv, H. Tenkanen, and E. Di Minin. Social media data for conservation science: A methodological overview. *Biological Conservation*, 233:298–315, 2019. ISSN 0006-3207. doi: <https://doi.org/10.1016/j.biocon.2019.01.023>. URL <https://www.sciencedirect.com/science/article/pii/S0006320718317609>.