**Instruções para a coleta de dados**

Esta pasta serve para hospedar os dados de coordenadas geográficas extraídas de artigos incluídos no topic-modeling que foram publicados antes do ano de 2000. Estes dados complementam aqueles extraídos do HTMLs dos artigos, cujo formato começou a ser disponibilizado em número razoável somente após o ano de 2000. Assim, vemos a necessidade de extrair manualmente a informação geográfica dos artigos antigos e digitalizados. Estes dados serão utilizados para o mapeamento do conhecimento em ciência recifal na costa brasileira.

A ideia é fazer o maior esforço possível de busca para extrair as coordenadas geográficas dos artigos. “Esforço” me refiro a pegar informação de texto, figuras, tabelas, gráficos, … de qualquer seção de um artigo. Portanto, o nome dos sítios de estudo também é um registro válido, bem como a informação de posição geográfica fornecida em uma figura (onde você pode visualmente coleta a informação geográfica).

Para sabermos exatamente a qualidade do dado sendo obtido, criei junto com a lista de artigos [planilha “Papers\_list\_used\_31\_03\_21[before2000].xlsx”] uma coluna chamado “Type\_of\_information”, onde você irá registrar o tipo de informação obtida: se foi “site\_name”, “coordinate\_from\_figure”,”coordinate\_from\_table”, e assim por diante.

Em seguida tem uma coluna chamada “site\_name”, para registrar o nome do sítio para aqueles estudos que fornecem somente esta informação.

Existe então a coluna “Position\_in\_the\_text”, onde deve registrar se foi “figure\_in\_results”, “table\_in\_results”, “table\_in\_maps”, “figure\_in\_supporting\_information”, e assim por diante.

Finalmente temos então as colunas de registro das coordenadas “Latitude” e “Longitude”. Percebam que existe “Latitude\_original” e “Longitude\_original”. Nestas colunas deves registrar a coordenada no formato original apresentado no artigo. Então, nas colunas “Latitude” e “Longitude” você irá registrar a coordenada no formato **“graus decimais”**. É bem possível que as coordenadas que vocês encontrarão estarão em outros formatos, então para evitar erros e perda de dados, sempre registrem a coordenada original em “Latitude\_original” e “Longitude\_original”. Para encontrar sítios e converter as coordenadas em graus decimais (quando necessário), vocês podem utilizar o Google Earth [nele é possível selecionar diferentes formatos de coordenada (guia: “Ferramentas”, “Opções”, “Mostrar lat/long”)].

No caso de um estudo ter mais de uma localidade (e.g., Abrolhos e Arraial do Cabo), ADICIONE linhas a mais logo abaixo da linha da referência, e adicione quantos sítios forem necessários.

Segue abaixo alguns exemplos de formato de coordenadas:

**Formato que deve ser:**

Decimal degrees (DD): 41.40338, 2.17403.

**Exemplos de outros formatos (que devem ser convertidos para DD):**

Degrees, minutes, and seconds (DMS): 41°24'12.2"N 2°10'26.5"E.

Degrees and decimal minutes (DMM): 41 24.2028, 2 10.4418.

Atentem que podem haver dúvidas devido a erros em registro de coordenadas. Registrem qualquer dúvida que tiverem neste arquivo mesmo (logo abaixo). Ficarei monitorando as dúvidas que vocês registrarem aqui.

Para registrar uma dúvida, sempre faça referência ao artigo sendo analisado (de acordo com PAPER\_ID do artigo, número da linha, constantes na planilha excel “Papers\_list\_used\_31\_03\_21[before2000]”) e, se possível, também o link para acessar o artigo.

André Luza

## Dúvidas

Elisa- Oi André, tudo certo! Posso começar com os outros artigos. Hanna, se precisar de ajuda me chama!

Hanna- Oi! Terminei também :)