Deciphering space-time jellyfish diversity blooms in the Mediterranean Sea through text mining

(Détermination de la diversité spatio-temporelle des méduses en Méditerranée grâce à la fouille de textes)

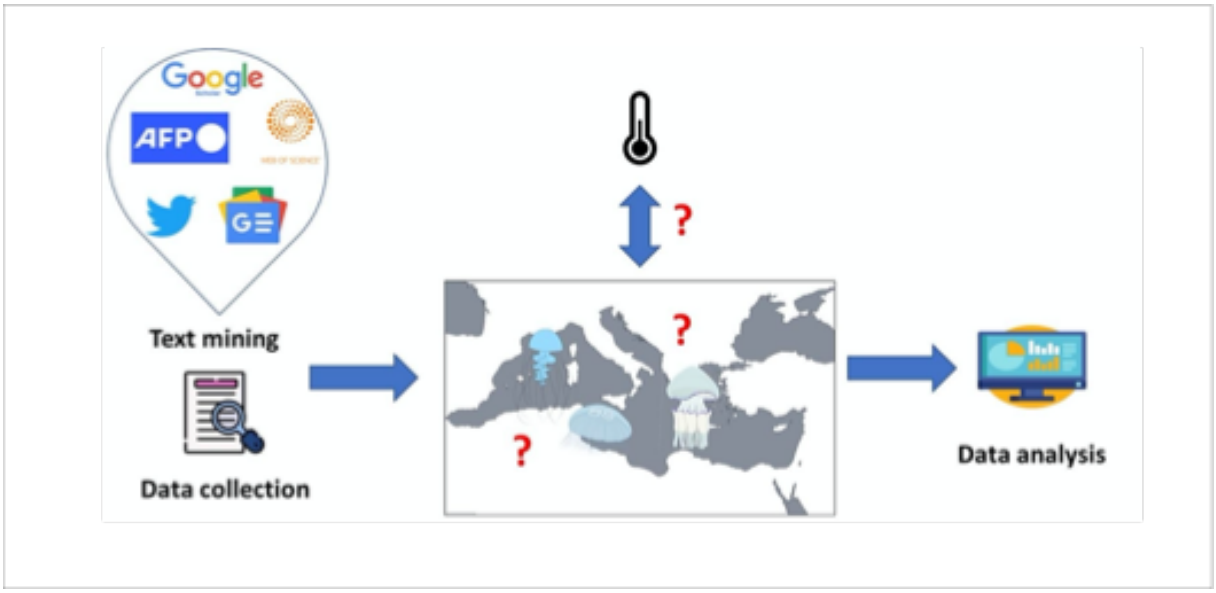
Context:

The Mediterranean Basin is a marine biodiversity hotspot exposed to increasing anthropogenic pressure which, combined with climate change, threatens the conservation of its biodiversity and the ecosystem services provided to riparian countries. In recent decades, massive blooms of jellyfish (here we consider stranded macroscopic gelatinous organisms) observed in temperate and subtropical regions have altered coastal ecosystems and economic activities, such as tourism and fishing.

The unpredictable nature of this phenomenon challenges traditional scientific approaches that have triggered multilateral research efforts and large-scale citizen science programs worldwide to assess the socio-ecological risks involved (e.g. the Mediterranean JellyWatch conducted by the International Commission for the Scientific Exploration of the Mediterranean Sea (www.ciesm.org)). Jellyfish blooms raise major ecological, economic, and even human health challenges, as some jellyfish species are venomous. In addition, global change scenarios such as warming, coastal habitat modification and increased shipping, warn on the enhancement of jellyfish blooms, as well as a reconfiguration in their basin-wide distribution, as these factors favor an increase in jellyfish abundance (Molinero et al. 2008; Lejeusne et al. 2010). Understanding biodiversity changes in these communities and their biogeography is therefore crucial for developing resilience management strategies in the already threatened Mediterranean ecosystems.

To date, most studies on biodiversity changes in the Mediterranean sea have focused either on species of economic interest (i.e. fish) or on vulnerable species (i.e. *Posidonia oceanica* seagrass). Jellyfish can be considered as suitable ecological indicators of the ecosystems state, since much of the observed variability can be explained with few descriptors (Leoni et al., 2021). Due to the countries' unequal research capacity or difficulty to access some areas in order to monitor jellyfish species, there is no doubt that scientifc monitoring cannot alone allow depicting their biodiversity changes at this basin scale. We consider that data collection from social networks and mass media could provide valuable data on spatial and temporal coverage contributing to the "big data" era of Ecology.

We here propose a complementary approach based on data gathered through text mining, which is the process of automatically deriving information from large amounts of text, and data mining to depict patterns and trends in collected data. Our goals are to map space-time patterns of stranded jellyfish in the Mediterranean basin, and to assess their diversity and the contribution of exotic species to such events. Finally, we aim at testing whether these events and patterns of exotic species are related with tropicalization and meridionalisation phenomena of the Mediterranean Sea.



The global architecture

Main tasks:
We will adapt the pipeline described in (Valentin, 2020) to the thematic of the project. The student's work will be structured in six steps:

1. Data collection: the student will collect news articles from Google News and AFP, posts from twitter and abstracts of scientific publications from Web of Science and Google Scholar. The queries will be based on a list of key words defined by the experts (e.g. jellyfish species' names, stranding, etc.). Changes in the Mediterranean climate over the past century will be assessed by using a climate proxy based on monthly anomaly fields of ocean- atmosphere variability (NCEP/NCAR Reanalysis dataset).
2. Data pre-treatment: We will obtain a multilingual corpus of documents (that covers coastal Mediterranean countries) indexed with meta-data (date, source...) which will be cleaned (removing of HTML tags, special characters...) and be structured in a database. All the documents will be translated in English as pivot language.
3. Document classification: the data classification step aims to select relevant documents, i.e., which reports an event related to the appearance of jellyfish (jellyfish overgrowth, prevention and control measures...) relevant for the expert. The classification module will use classical supervised machine learning approach. To initiate the process, the experts will have to validate a significant number of documents to obtain sufficient labeled data.
4. Extraction of terms: the student will extract entities such as location (GeoNames gazetteer), date (HeidelTime system), and thematic entities based on a list of words given by the experts which will be extend automatically by using dictionaries.
5. Finally, the extracted knowledge will be available in a visual analytics tool that helps to visualize and explore data for monitoring jellyfish overgrowth (Goel, 2020).
6. This information will be considered in the context of tropicalisation of the Mediterranean Sea.

References:

1. Goel R., Valentin S., Delaforge A., Fadloun S., Sallaberry A., Roche M., Poncelet P. (2020). EpidNews: Extracting, exploring and annotating news for monitoring animal diseases. J. Comput. Lang. 56: 100936 (2020)
2. Lejeusne C., Chevaldonné P., Pergent-Martini C., Boudouresque C., Perez T. (2010). Climate change effects on a miniature ocean: The highly diverse, highly impacted Mediterranean Sea. Trends in Ecology and Evolution, 25, 250-260.
3. Leoni V, Bonnet D, Ramírez Romero E, Molinero JC (2021). Biogeography and phenology of the jellyfish *Rhizostoma pulmo* (Cnidaria: Scyphozoa) in southern European Seas. Journal of Global Ecology and Biogeography 30:622–639. https://doi.org/10.1111/geb.13241.
4. Molinero JC, Casini M, Buecher E (2008). The influence of the Atlantic and regional climate variability on the longterm changes in gelatinous carnivore populations in the northwestern Mediterranean. Limnol Oceanogr 53:1456-1467.
5. Valentin S., Arsevska E., Falala S., de Goër J., Lancelot R., Mercier A., Rabatel J., Roche M. (2020). PADI-web: A multilingual event-based surveillance system for monitoring animal infectious diseases, Computers and Electronics in Agriculture, Volume 169, 105163, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2019.105163.

Application: Applications for this position (CV, Motivation Letter, last grade report, References) will be received EXCLUSIVELY in a single PDF document accessible for download via email sent to Sandra Bringay (Sandra.Bringay@lirmm.fr) and Delphine BONNET (delphine.bonnet@umontpellier.fr).

Supervision :

- UMR 5506 LIRMM Equipe ADVANSE : BRINGAY Sandra, SALLABERRY Arnaud, PONCELET Pascal
- UMR 9190 MARBEC Equipe Etat des lieux de la biodiversité/Fonctionnement et évolution des organismes et des écosystèmes : BONNET Delphine, MOLINERO Juan Carlos

Location : LIRMM et MARBEC- Campus Triolet

Gratification : 6 month