Natural language processing and an automated trait-mining workflows for studying plant size distributions on islands

Dr. Helena Bestová

Summary

Plant size is related to many key ecosystem properties such as carbon storage capacity and productivity. Recently, the development of extensive biological databases enabled ecologists to incorporate plant size in biodiversity models. However, these databases contain a lot of missing data, even though the missing information can be found in thousands of scatter resources such as Floras, monographs and scientific papers. To harness this information for macroecological research, I will develop and apply cutting-edge machine learning models for natural language processing and an automated trait-mining workflow. This will allow me to mobilize previously neglected information from existing literature and fill gaps in global trait databases. I will use the collected data to develop a model for predicting size distributions of plants on islands. The model will improve our knowledge of the mechanistic drivers of plant size distributions, advancing our understanding of ecosystem functioning.