**Book Recommendations from Charles Darwin**

Build a book recommendation system using NLP and the text of books like "On the Origin of Species."

#### Project Description

Recommendation systems are at the heart of many products such as Netflix or Amazon. They generally rely on metadata (e.g., the actors or director of a movie) or on user tastes (e.g., the movies you liked before) to determine which you are most likely to enjoy. But when you are working with text-heavy datasets, you have access to a much richer resource—the whole text! In this project, you will learn how to build the basis of a book recommendation system based on their content. You will use Charles Darwin's bibliography to find out which books might interest you.

To complete this project, you should be familiar with the basics of Python, Pandas and Natural Language Processing. [pandas Foundations](https://www.datacamp.com/courses/pandas-foundations) and [Natural Language Processing Fundamentals in Python](https://www.datacamp.com/courses/natural-language-processing-fundamentals-in-python) are recommended as prerequisites.

The dataset was manually collected from [Project Gutenberg](https://www.gutenberg.org).

#### Project Tasks

* 1 Darwin's bibliography
* 2 Load the contents of each book into Python
* 3 Find "On the Origin of Species"
* 4 Tokenize the corpus
* 5 Stemming of the tokenized corpus
* 6 Building a bag-of-words model
* 7 The most common words of a given book
* 8 Build a tf-idf model
* 9 The results of the tf-idf model
* 10 Compute distance between texts
* 11 The book most similar to "On the Origin of Species"
* 12 Which books have similar content?