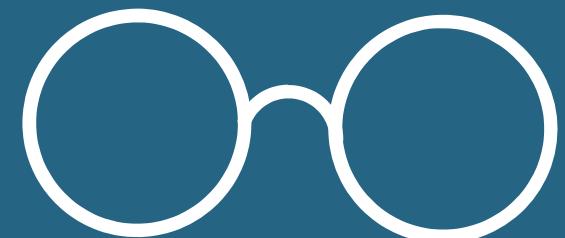


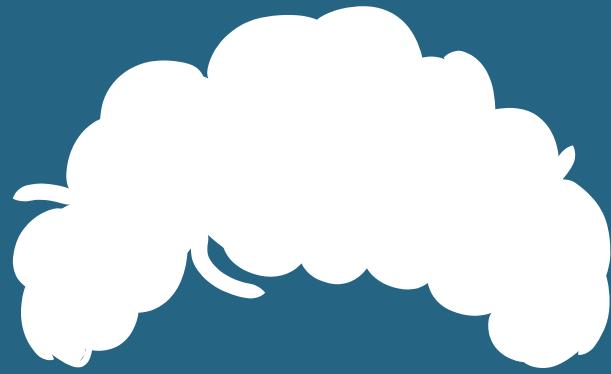
SEA ME LIKE I SEE YOU: SEA CREATURES SPECIES SEARCH ENGINE

PROPOSAL PRESENTATION

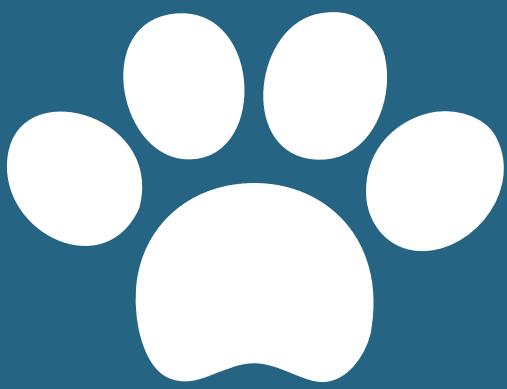
Team Members



Waris Damkham
6388014



Thanawath Huayhongthong
6388016



Pattanan Korkiattrakool
6388022



WHAT IS YOUR SEARCH SYSTEM FOR?

- ~ Learning marine life
- ~ Generate a list of marine life
- ~ Develop a straightforward approach for marine life-curious reader



WHAT ARE YOUR TARGET USER GROUP?

DIVIDED INTO 3 MAIN GROUP

- sea creatures enthusiasts
- Scuba enthusiasts 
- Divers 
- Also students and teachers for educational purpose



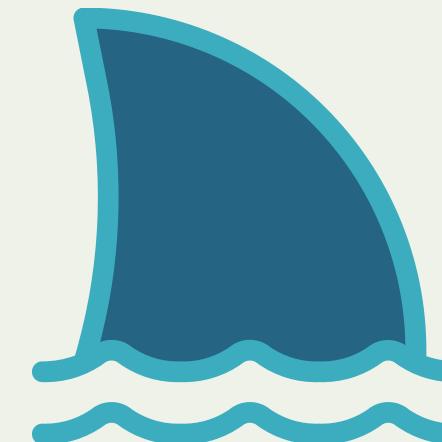
WHAT IS YOUR DATA SOURCE?

MARINEBIO

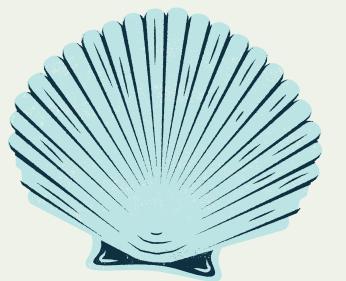
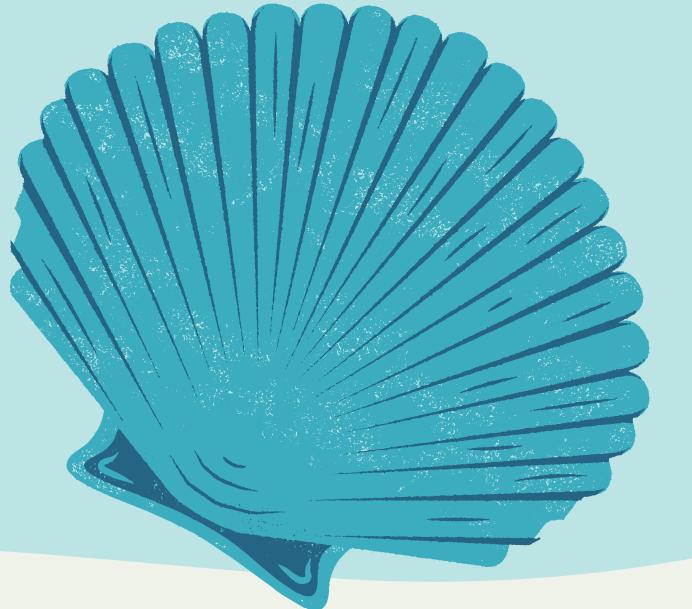
FISHBASE

WHAT & WHY MARINEBIO?

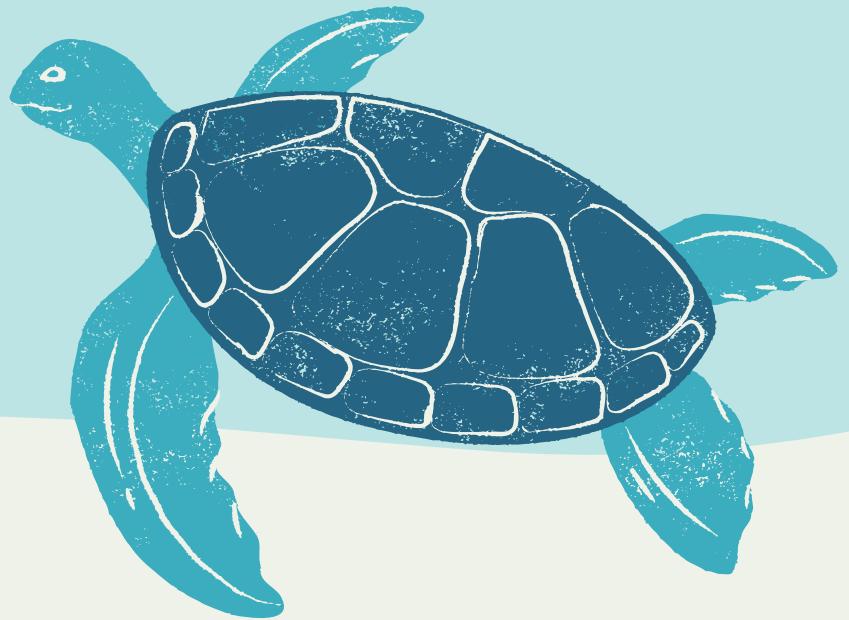
OPEN SOURCE DATA



CONTAIN RELEVANT
CONTENT



WHAT IS FISHBASE?



GLOBAL ONLINE DATABASE



That provides key information on the taxonomy,
distribution, and ecology of all marine species in the world .



WHY FISHBASE?

COLLABORATIVE OPEN SOURCE DATA

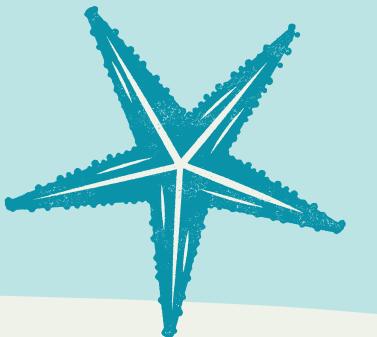
anyone may use it
without cost.

CONTAIN RELEVANT CONTENT

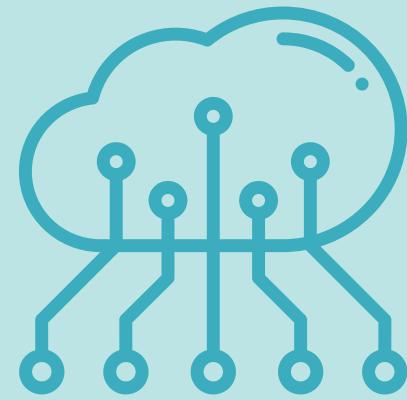
suitable for our group
topic.

PROVIDE APIs

fetch data and crawl
through their website
contents.



HOW TO PREPROCESS YOUR DATA?



Collect data from data sources to create and store data in Dataset in an external file.

abc

Keep order in alphabetical order

WE USING 3 TECHNIQUES

STOP WORD

- cut out unnecessary words
- “The great white shark”, so when we remove stop word, it will be “great white shark”.

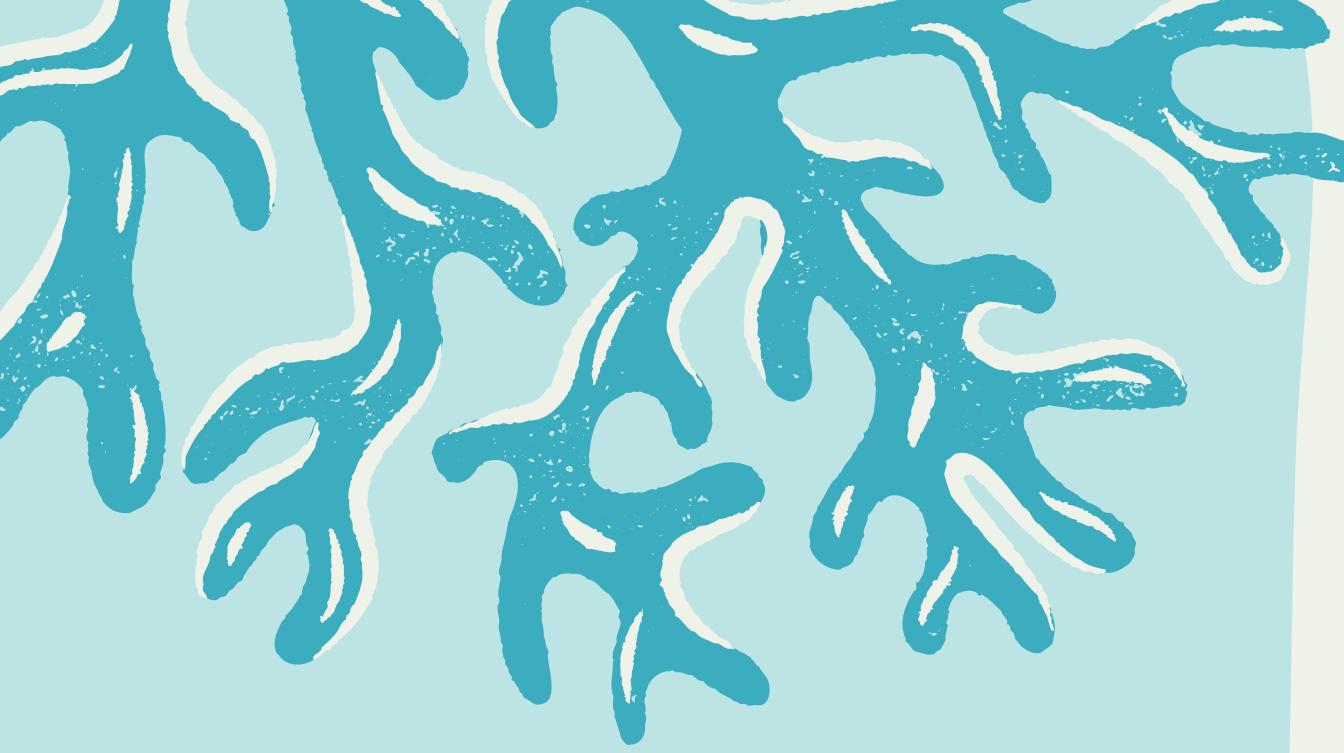
STEMMING

- chop off the ends of words
- the word ‘whity’ back into an original form, so the word will be reverted as “white”.

TOKENIZATION

- splitting input textual data into separate meaningful pieces
- “Great White Shark” will be separated into 3 tokens which are |Great|, |White|, |Shark|.

HOW TO INDEX THE DOCUMENTS



K-GRAM

- K-gram is the division of a word proportion of 1 word 3 letters called a trigram.
- In this way it is suitable for our search engine because our data is very large. Therefore, by dividing that word, the search time will be faster and the closest other results can be found.

HOW TO RANK THE DOCUMENTS

SORTS
ALPHABETICALLY

SORTED BY
NATION

THANK YOU

