

August 2023

Chatbots About Coral Reef Fish



Steve Desilets

Agenda

1



Introduction

2



Problem Definition

3



Methodology

4



Results

5



Conclusions and Recommendations

6



References

7



Appendix

8



Questions

Introduction

Context That Inspired Chatbots About Coral Reef Fish

Over Winter Break, I obtained my Open Water scuba diving certification. While I loved the experience, I quickly realized that I didn't know anything about the marine life I was observing, including:

- How to recognize species based on their appearances
- The ways in which different species might be dangerous
- How humans should engage with different species

Scuba divers should be familiar coral reef fish for their own safety. This knowledge can also help divers understand how best to engage with coral reef fish, which represent 25% of marine fish species despite the many challenges they face.

For this course's final project, I set out to create chatbots that could help educate scuba divers about coral reef fish.



Problem Definition

Below, we provide the objectives that we set out to achieve when constructing chatbots about coral reef fish.



Help Divers Recognize Coral Reef Fish

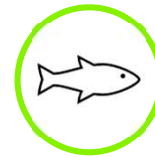
These chatbots should help educate divers to recognize coral reef inhabitants based on visual descriptions and to answer divers' questions about the behaviors and characteristics of those species.



Advance the Safety of Scuba Divers

These chatbots should be able to educate divers about the many mechanisms that coral reef inhabitants leverage to defend themselves and to hunt, including:

- Biting
- Stinging
- Venom / Poison
- Electrocutation
- Ambush Predation
- Parasitic Attachment



Support the Welfare of Coral Reef Fish

Ideally, these chatbots should be able to educate divers about the array of challenges that coral reef fish face, including predation, habitat loss and degradation, pollution, and overfishing – many of which are caused by humans.

Methodology

This slide summarizes the methodology leveraged to construct each of the three initial chatbots focused on coral reef fish. Notably, the corpus leveraged for these chatbots is Wikipedia's "Coral reef fish" webpage (Wikipedia 2023). Given that the resulting document contained 5,550 words, we were confident that the chatbot would have a rich corpus from which to extract information about coral reef fish.

1

Chatbot 1 – Sentence Transformer Chatbot

This pre-trained Sentence Transformer Model, which was pre-trained via 215 million question-answer pairs, encodes each sentence of the corpus, as well as the user's question, as 384-dimensional vectors. Then, the model calculates the cosine similarity between the question vector and each of the corpus sentence vectors and returns the corpus sentence associated with the highest cosine similarity with the user's question (Espejel 2022).

2

Chatbot 2 – OpenAI ChatGPT-2-Based Chatbot

This chatbot was created by fine tuning the ChatGPT-2 model in Python using sample training questions about coral reef fish that were entirely separate from our testing questions (OpenAI 2019). After finishing this fine-tuning process, this large language model encoded and responded to each user question.

3

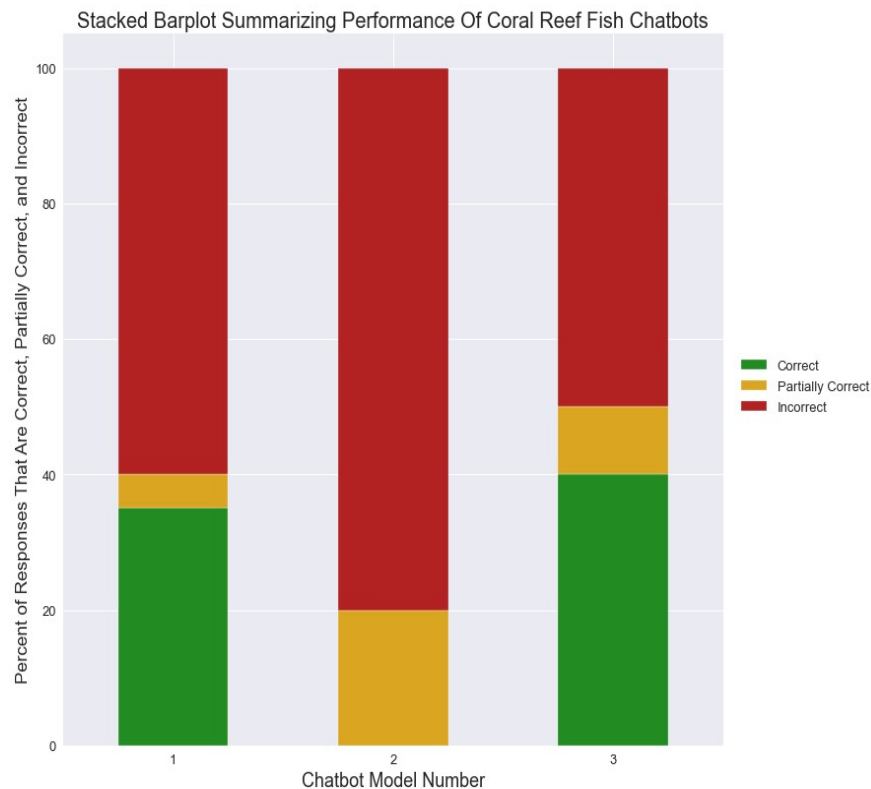
Chatbot 3 – Basic TF-IDF-Based Chatbot

This chatbot first tokenizes the corpus by sentence. It then conducts data wrangling on each sentence in the corpus and in the user's question - including word tokenization, lemmatization, transformation to lowercase, and punctuation removal. The chatbot applies TF-IDF vectorization to each sentence, calculates the cosine similarity between each of the sentences, and returns the sentence from the corpus with the highest cosine similarity to the user's question (Kulkarni 2020).

Results – Chatbot Performance Metrics

The stacked bar plot and key takeaways below visualize and summarize performance metrics related to the tests of each of our chatbots with 20 questions that are provided in the appendix.

Visualization of Chatbot Performance



Key Takeaways

40%

Correct or Partially Correct Answers from Chatbot 1
(the Sentence Transformer Chatbot)

20%

Correct or Partially Correct Answers from Chatbot 2
(the Fine-Tuned ChatGPT-2-Based Chatbot)

50%

Correct or Partially Correct Answers from Chatbot 3
(the Basic TF-IDF-Based Chatbot)

Results – Chatbot Response Examples and Key Takeaways

The table below displays the chatbot responses to 5 of the 20 test questions. The table's cells are color coded as green, yellow, and red to reflect whether each answer was correct, partially correct, or incorrect, respectively.

Question / Chatbot Response	Question 7 and Responses	Question 9 and Responses	Question 12 and Responses	Question 14 and Responses	Question 17 and Responses
Question	What species are parasitic?	What species are known for attacking scuba divers?	How do fish get rid of their parasites?	Which species have a commensalistic relationship?	Which species are ambush predators?
Chatbot 1 Response	Monogenean parasites of the genus pseudorhabdosynochus (arrows) on the gill filament of a grouper	They are swift predators who patrol the reef in hunting packs	Some fish specialize as cleaner fish, and establish cleaning stations where other fish can come to have their parasites nibbled away	The relationship can be mutualistic, when both species benefit from the relationship, commensalistic, when one species benefits and the other is unaffected, and parasitic, when one species benefits, and the other is harmed.	Another ambush predator is the striated frogfish (right).
Chatbot 2 Response	Sharks and giant moray	Sharks and giant moray	Fish eat parasites, and fish eat fish. Fish eat parasites because fish eat parasites. Fish eat fish because fish eat parasites. Fish eat fish because fish eat parasites. Fish eat fish because fish eat parasites. Fish eat fish because fish eat parasites. Fish eat fish because fish eat parasites. Fish eat fish because fish eat parasites. Fish eat fish because fish eat parasites. Fish eat fish because fish eat parasites.	Sharks and giant moray	Sharks and giant moray
Chatbot 3 Response	Parasites of coral reef fish include nematodes, platyhelminthes (cestodes, digeneans, and monogeneans), leeches, parasitic crustaceans, such as isopods and copepods, and various microorganisms such as myxosporidia and microsporidia	Lionfish can aggressively dart at scuba divers and attempt to puncture the facemask with their venomous spines	As with all fish, coral reef fish harbour parasites	The relationship can be mutualistic, when both species benefit from the relationship, commensalistic, when one species benefits and the other is unaffected, and parasitic, when one species benefits, and the other is harmed.	The well camouflaged striated frogfish, a species of anglerfish is an ambush predator

Findings

- + Both this slide and the previous slide underscore that chatbots 1 and 3 performed reasonably well - though both chatbots have significant room for improvement.
- + The Chatbot 1 and 3 responses to questions 9 and 14 reflect that sometimes the chatbots focus too much on the wrong keywords (i.e. “attacking” and “relationships”) rather than some other keywords that would have been helpful (i.e. “scuba divers” and “species”).
- + Chatbot 2 performed relatively poorly – often producing rambling, semi-coherent text or simply providing the phrase “sharks and giant moray”.

Conclusions and Recommendations

Below, we summarize the primary conclusions and recommendations resulting from this exercise in constructing chatbots focused on coral reef fish.

Conclusions

Chatbots Can Help Educate Divers

While none of the chatbots performed perfectly, some proved moderately useful in providing information about coral reef fish to divers. Furthermore, when the chatbots answered incorrectly, it was almost always because they returned irrelevant – not factually incorrect – answers. Therefore, these chatbots (especially chatbots 1 and 3) seem to pose little risk of generating fake and misleading information.

Chatbot Performance Varied Widely

The performance of these three models varied quite widely. Perhaps the best evidence of the wide range of performance metrics is that the OpenAI-based chatbot answered 0% of the test questions fully correctly, while the Basic TF-IDF-Based Chatbot answered 40% of the test questions fully correctly.

Recommendations

Try More Chatbot Development Methods

The researchers should experiment with additional approaches to developing chatbots, exploring generative Question-Answering Chatbots available on Hugging Face.

Experiment How To Improve Existing Chatbots

The research team should experiment with strategies to improve the performance of the existing chatbots, such as leveraging different data wrangling / vectorization methods for the first and third chatbots and attempting additional fine-tuning of the OpenAI-based chatbot via more training questions.

Implementation

If a client could not wait for further experimentation to improve these chatbots prior to implementation, then I would recommend implementing the third, Basic TF-IDF-Based Chatbot, which performed most strongly in answering the test questions and which poses no risk of generating false information.

References

- + Espejel, Omar. 2022. “multi-qa-MiniLM-L6-cos-v1.” *Hugging Face*. <https://huggingface.co/sentence-transformers/multi-qa-MiniLM-L6-cos-v1>.
 - + Kulkarni, Mandar. 2020. “NLP Chatbot using nltk.” *Github*. <https://github.com/mandar196/InstaBot/blob/master/NLP%20Chatbot%20usnig%20nltk.ipynb>.
 - + OpenAI. 2019. “Better language models and their implications.” *Hugging Face*. <https://openai.com/research/better-language-models>.
 - + Wikipedia. 2023. “Coral Reef Fish” *Wikipedia*. https://en.wikipedia.org/wiki/Coral_reef_fish.
-

Appendix – Questions Leveraged To Test Chatbot Performance

The table below summarizes the first 12 of the 20 questions used to test the performance of each of the chatbots, as well as the target answers derived from the corpus of information about coral reef fish.

Question Number	Questions	Target Answer From Corpus
1	How many species of fish live in coral reefs?	6,000 – 8,000 species
2	What is the most venomous fish?	Reef stonefish
3	What sharks live in coral reefs?	Whitetip, blacktip, and grey reef sharks (which are all species of requiem shark), as well as the Galapagos shark, the tawny nurse shark, and hammerheads
4	What fish are poisonous?	Giant moray and spotted trunkfish
5	What fish can electrocute you?	Stargazer
6	What species can sting you?	Fire coral, stingrays, and sea anemones
7	What species are parasitic?	Nematodes, Platyhelminthes (cestodes, digeneans, and monogeneans), leeches, parasitic crustaceans such as isopods and copepods, and various microorganisms (such as myxosporidia and microsporidia)
8	What species are venomous?	Spinefoot rabbitfish, coral rabbitfish, reef stonefish, stargazer, lionfish
9	What species are known for attacking scuba divers?	Lionfish can aggressively dart at scuba divers and attempt to puncture the facemask with their venomous spines
10	What are common herbivorous fish?	The four largest groups of coral reef fishes that feed on plants are the parrotfishes, damselfishes, rabbitfishes, and surgeonfishes
11	Why do fish camouflage themselves?	Ambush predation or predator avoidance
12	How do fish get rid of their parasites?	Some fish specialize as cleaner fish and establish cleaning stations where other fish can come to have their parasites nibbled away

Appendix – Questions Leveraged To Test Chatbot Performance

The table below summarizes the last 8 of the 20 questions used to test the performance of each of the chatbots, as well as the target answers derived from the corpus of information about coral reef fish.

Question Number	Questions	Target Answer From Corpus
13	What species have mutualistic relationships?	Clownfish and sea anemones
14	What species have commensalistic relationships?	Hawkfish and fire coral Or Pearlfish and sea cucumber
15	Where in the world are coral reefs found?	The Indo-Pacific and the Tropical Western Atlantic
16	What species engage in schooling?	Parrotfish, surgeonfish, barracuda, spinefoot rabbitfish, Yellowfin goatfish, blue-striped snapper, grouper, cardinalfish, and bigeye trevally
17	Which species are ambush predators?	Reef stonefish, tasseled scorpionfish, giant moray, and striated frogfish
18	Why are some coral reef fish colorful?	Recognition during mating, warning predators of poison or venom, or confusing predators
19	Which species is blue?	Synchiropus splendidus
20	What is the slowest species that lives in coral reefs?	Sea horses

Questions?