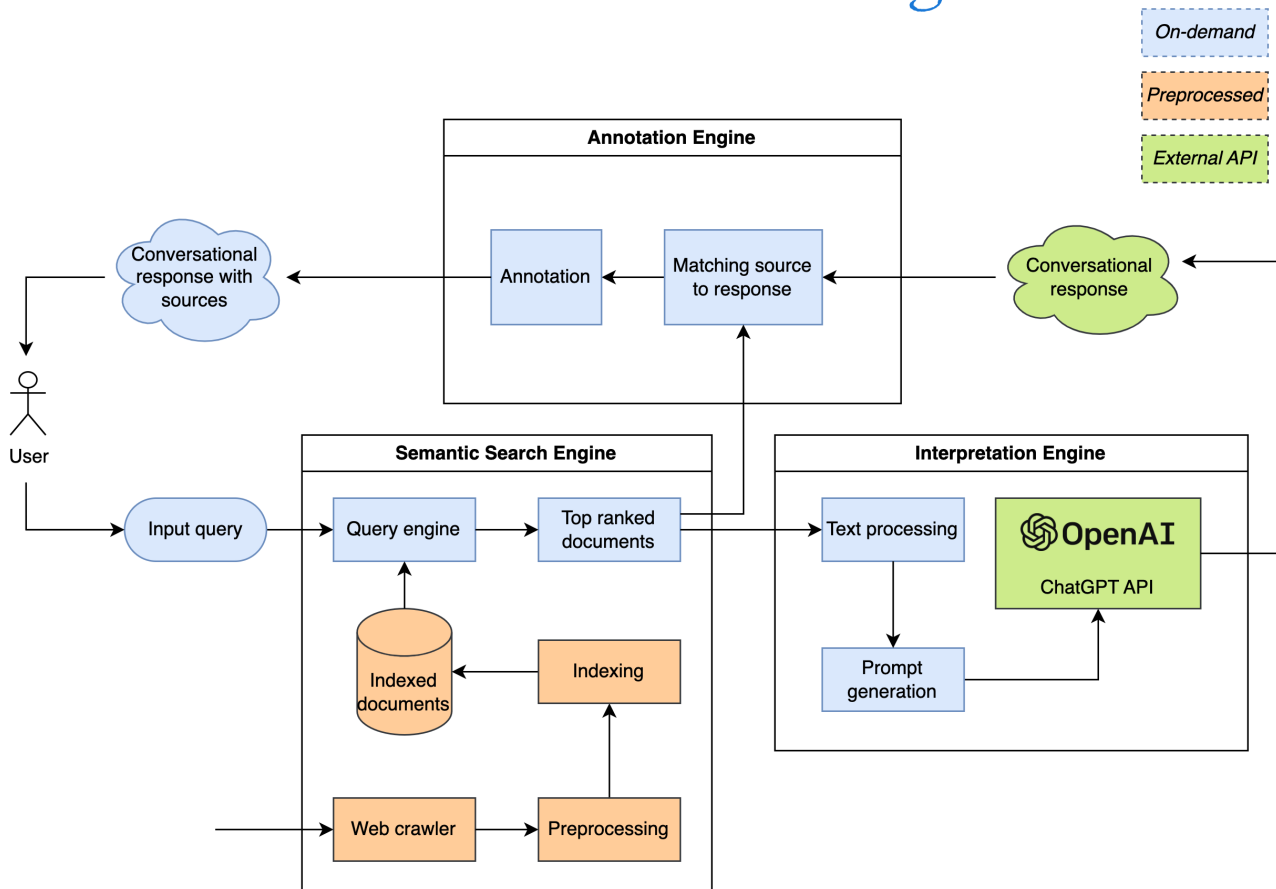


On February 7, 2023, Microsoft launched a new version of Bing that included ChatGPT to enhance search results. Inspired by this advancement and the movie "Avatar: The Way of Water," I created Cetasearch, a search engine that focuses on knowledge about the ocean and cetacean species such as whales, orcas, dolphins, and more. Cetasearch offers conversational answers with detailed links to the sources, which I hope could play a small role in the conservation of the ocean and marine wildlife.

The search engine has three main functionalities: a semantic search engine, generative text completion for conversational answers, and annotations that match answers with sources. Everything was built using publicly available NLP pre-trained models and APIs, such as Sentence transformer, OpenAI's GPT3.5 API, and multiple pre-trained LLMs during development.

## Cetasearch Architecture Diagram



**|** *Data crawling, text processing and indexing is pre-calculated in batch to save inference time*

Due to my limited computational power of my computer, multiple measures were taken to narrow the scope of this project while maintaining the core functionalities:

- Focus on the ocean and cetacean species. The data sources were limited to a small subset of Wikipedia articles centered around these topics.
- Preprocess and index the source documents to save processing time.
- Leverage OpenAI's GPT3.5 API for text completion. During the experimentation phase, I tried a few open-source LLM models as well, but OpenAI's API provided the best answers by a large margin.

Ask me whale-ever you want...

**Question:**

How can I help protecting whales?

**Answer:**

There are several ways to help protect whales[0]. You can support organizations that work to protect whales, such as the International Whaling Commission, the Marine Mammal Protection Act of 1972, and the World Wildlife Fund. You can also support legislation that protects whales, such as the Endangered Species Act[1]. You can also help reduce marine pollution by reducing your use of single-use plastics and other pollutants. You can also help reduce the risk of whales being caught in fishing nets by supporting sustainable fishing practices. Finally, you can help raise awareness of the threats facing whales by sharing information about them with your friends and family[0].

**Sources****[0] Whale (Wikipedia)**

Other threats. Whales can also be threatened by humans more indirectly. They are unintentionally caught in fishing nets by commercial fisheries as bycatch and accidentally swallow fishing hooks. Gillnetting and Seine netting is a significant cause of mortality in whales and other marine mammals. Species commonly entangled include beaked whales. Whales are also affected by marine pollution. High levels of organic chemicals accumulate in these animals since they are high in the food chain. They have large reserves of blubber, more so for toothed whales as they are higher up the food chain than baleen whales. Lactating mothers can pass the toxins on to their young. These pollutants can cause gastrointestinal cancers and greater vulnerability to infectious diseases. They can also be poisoned by swallowing litter, such as plastic bags. Advanced military sonar harms whales. Sonar interferes with the basic biological functions of whales—such as feeding and mating—by impacting their ability to echolocate. Whales swim in response to sonar and sometimes experience decompression sickness due to rapid changes in depth. Mass strandings have been triggered by sonar activity, resulting in injury or death. Whales are sometimes killed or injured during collisions with ships or boats. This is considered to be a significant threat to vulnerable whale populations such as the North Atlantic right whale, whose total population numbers less than 500.

**[1] Cetacea (Wikipedia)**

Whaling. Whaling is the practice of hunting whales, mainly baleen and sperm whales. This activity has gone on since the Stone Age. In the Middle Ages, reasons for whaling included their meat, oil usable as fuel and the jawbone, which was used in house construction. At the end of the Middle Ages, early whaling fleets aimed at baleen whales, such as bowheads. In the 16th and 17th centuries, the Dutch fleet had about 300 whaling ships with 18,000 crewmen. In the 18th and 19th centuries, baleen whales especially were hunted for their baleen, which was used as a replacement for wood, or in products requiring strength and flexibility such as corsets and crinoline skirts. In addition, the spermaceti found in the sperm whale was used as a machine lubricant and the ambergris as a material for pharmaceutical and perfume industries. In the second half of the 19th century, the explosive harpoon was invented, leading to a massive increase in the catch size. Large ships were used as "mother" ships for the whale handlers. In the first half of the 20th century, whales were of great importance as a supplier of raw materials. Whales were intensively hunted during this time; in the 1930s, 30,000 whales were killed. This increased to over 40,000 animals per year up to the 1960s, when stocks of large baleen whales collapsed. Most hunted whales are now threatened, with some great whale populations exploited to the brink of extinction. Atlantic and Korean gray whale populations were completely eradicated and the North Atlantic right whale population fell to some 300–600. The blue whale population is estimated to be around 14,000. The first efforts to protect whales came in 1...

Cetasearch provide a better answer to a generic answer than Google's current feature snippet!

how can I help protecting whales?

×

🔊

📷

🔍

🔍 All

🖼️ Images

📺 Videos

📰 News

🛒 Shopping

⋮ More

Tools

About 24,900,000 results (0.55 seconds)

5 Simple Ways to Help Protect Whales


1. Keep an eye on your carbon footprint.

2. Use fewer plastic-made products.


3. 3: Support organizations working to protect the ocean.

4. Travel responsibly.

5. 5: Influence change in your community.



5 SIMPLE WAYS TO HELP PROTECT WHALES



The Poor Traveler

https://www.thepoortraveler.net › 2015/08 › humpback-...

5 Simple Ways to Help Protect Whales - The Poor Traveler

?

 About featured snippets
 

!

 Feedback

Other open-ended questions got good answers as well.

Ask me whale-ever you want...

**Question:**

Tell me five interesting facts about dolphin

**Answer:**

1. Dolphins are highly social animals, often living in pods of up to a dozen individuals[0].
2. They establish strong social bonds, and will stay with injured or ill members, helping them to breathe by bringing them to the surface if needed[0].
3. Dolphins communicate using a variety of clicks, whistle-like sounds and other vocalizations[2].
4. Dolphins display culture, something long believed to be unique to humans[0].
5. Male dolphins can get into disputes over companions and females[0].

**Sources****[0] Dolphin (Wikipedia)**

*Behavior. Socialization. Dolphins are highly social animals, often living in pods of up to a dozen individuals, though pod sizes and structures vary greatly between species and locations. In places with a high abundance of food, pods can merge temporarily, forming a superpod; such groupings may exceed 1,000 dolphins. Membership in pods is not rigid; interchange is common. They establish strong social bonds, and will stay with injured or ill members, helping them to breathe by bringing them to the surface if needed. This altruism does not appear to be limited to their own species. The dolphin Moko in New Zealand has been observed guiding a female pygmy sperm whale together with her calf out of shallow water where they had stranded several times. They have also been seen protecting swimmers from sharks by swimming circles around the swimmers or charging the sharks to make them go away. Dolphins communicate using a variety of clicks, whistle-like sounds and other vocalizations. Dolphins also use nonverbal communication by means of touch and posturing. Dolphins also display culture, something long believed to be unique to humans (and possibly other primate species). In May 2005, a discovery in Australia found Indo-Pacific bottlenose dolphins (Tursiops aduncus) teaching their young to use tools. They cover their snouts with sponges to protect them while foraging. This knowledge is mostly transferred by mothers to daughters, unlike simian primates, where knowledge is generally passed on to both sexes. Using sponges as mouth protection is a learned behavior. Another learned behavior was discovered among river dolphins in Brazil, where some male dolphins use w...*

**[2] Dolphin (Wikipedia)**

A dolphin is an aquatic mammal within the infraorder Cetacea. Dolphin species belong to the families Delphinidae (the oceanic dolphins), Platanistidae (the Indian river dolphins), Iniidae (the New World river dolphins), Pontoporiidae (the brackish dolphins), and the extinct Lipotidae (baiji or Chinese river dolphin). There are 40 extant species named as dolphins. Dolphins range in size from the and Maui's dolphin to the and orca. Various species of dolphins exhibit sexual dimorphism where the males are larger than females. They have streamlined bodies and two limbs that are modified into flippers. Though not quite as flexible as seals, some dolphins can briefly travel at speeds of per hour or leap about . Dolphins use their conical teeth to capture fast-moving prey. They have well-developed hearing which is adapted for both air and water. It is so well developed that some can survive even if they are blind. Some species are well adapted for diving to great depths. They have a layer of fat, or blubber, under the skin to keep warm in the cold water. Dolphins are widespread. Most species prefer the warm waters of the tropic zones, but some, such as the right whale dolphin, prefer colder climates. Dolphins feed largely on fish and squid, but a few, such as the orca, feed on large mammals such as seals. Male dolphins typically mate with multiple females every year, but females only mate every two to three years. Calves are typically born in the spring and summer months and females bear all the responsibility for raising them. Mothers of some species fast and nurse their young for a relatively long period of time. Dolphins produce a variety of vocalizations,...

Ask me whale-ever you want...

**Question:**

Compare the intelligence of Blue whale, Orca, and Dolphin

**Answer:**

Blue whales have the largest brain mass of any animal on Earth, averaging and in mature males, in comparison to the average human brain which averages in mature males. Orcas have the second largest brain mass of any animal on Earth, next to the sperm whale[2]. The brain to body mass ratio in some is second only to humans. Dolphins are known to teach, learn, cooperate, scheme, and grieve. The neocortex of many species is home to elongated spindle neurons that, prior to 2007, were known only in hominids. In humans, these cells are involved in social conduct, emotions, judgment, and theory of mind. Cetacean spindle neurons are found in areas of the brain that are homologous to where they perform a similar function[0]. Allometric analysis indicates that mammalian brain size scales at approximately the 23 or 34 exponent of the body mass. Comparison of a particular animal's brain size with the expected brain size based on such allometric analysis provides an encephalization quotient that can be used as another indication of animal intelligence[1]. Therefore, Blue whales have the largest brain mass, followed by Orcas, and then Dolphins[2].

**Sources****[0] Dolphin (Wikipedia)**

*Intelligence. Dolphins are known to teach, learn, cooperate, scheme, and grieve. The neocortex of many species is home to elongated spindle neurons that, prior to 2007, were known only in hominids. In humans, these cells are involved in social conduct, emotions, judgment, and theory of mind. Cetacean spindle neurons are found in areas of the brain that are homologous to where they perform a similar function. Brain size was previously considered a major indicator of the intelligence of an animal. Since most of the brain is used for maintaining bodily functions, greater ratios of brain to body mass may increase the amount of brain mass available for more complex cognitive tasks. Allometric analysis indicates that mammalian brain size scales at approximately the 23 or 34 exponent of the body mass. Comparison of a particular animal's brain size with the expected brain size based on such allometric analysis provides an encephalization quotient that can be used as another indication of animal intelligence. Orcas have the second largest brain mass of any animal on earth, next to the sperm whale. The brain to body mass ratio in some is second only to humans. Self-awareness is seen, by some, to be a sign of highly developed, abstract thinking. Self-awareness, though not well-defined scientifically, is believed to be the precursor to more advanced processes like meta-cognitive reasoning (thinking about thinking) that are typical of humans. Research in this field has suggested that cetaceans, among others, possess self-awareness.*

**[1] Mammal (Wikipedia)**

*Intelligence. In intelligent mammals, such as primates, the cerebrum is larger relative to the rest of the brain. Intelligence itself is not easy to define, but indications of intelligence include the ability to learn, matched with behavioral flexibility. Rats, for example, are considered to be highly intelligent, as they can learn and perform new tasks, an ability that may be important when they first colonize a fresh habitat. In some mammals, food gathering appears to be related to intelligence: a deer feeding on plants has a brain smaller than a cat, which must think to outwit its prey. Tool use by animals may indicate different levels of learning and cognition. The sea otter uses rocks as essential and regular parts of its foraging behaviour (smashing abalone from rocks or breaking open shells), with some populations spending 21% of their time making tools. Other tool use, such as chimpanzees using twigs to "fish" for termites, may be developed by watching others use tools and may even be a true example of animal teaching. Tools may even be used in solving puzzles in which the animal appears to experience a "Eureka moment". Other mammals that do not use tools, such as dogs, can also experience a Eureka moment. Brain size was previously considered a major indicator of the intelligence of an animal. Since most of the brain is used for maintaining bodily functions, greater ratios of brain to body mass may increase the amount of brain mass available for more complex cognitive tasks. Allometric analysis indicates that mammalian brain size scales at approximately the or exponent of the body mass. Comparison of a particular animal's brain size with the ex...*

**[2] Whale (Wikipedia)**

*Intelligence. Whales are known to teach, learn, cooperate, scheme, and grieve. The neocortex of many species of whale is home to elongated spindle neurons that, prior to 2007, were known only in hominids. In humans, these cells are involved in social conduct, emotions, judgement, and theory of mind. Whale spindle neurons are found in areas of the brain that are homologous to where they are found in humans, suggesting that they perform a similar function. Brain size was previously considered a major indicator of the intelligence of an animal. Since most of the brain is used for maintaining bodily functions, greater ratios of brain to body mass may increase the amount of brain mass available for more complex cognitive tasks. Allometric analysis indicates that mammalian brain size scales at approximately the 23 or 34 exponent of the body mass. Comparison of a particular animal's brain size with the expected brain size based on such allometric analysis provides an encephalisation quotient that can be used as another indication of animal intelligence. Sperm whales have the largest brain mass of any animal on Earth, averaging and in mature males, in comparison to the average human brain which averages in mature males. The brain to body mass ratio in some odontocetes, such as belugas and narwhals, is second only to humans. Small whales are known to engage in complex play behaviour, which includes such things as producing stable underwater toroidal air-core vortex rings or "bubble rings". There are two main methods of bubble ring production: rapid puffing of a burst of air into the water and allowing it to rise to the surface, forming a ring, or swimming rep...*

# Features

## Currently available features

The following features are currently available:

- A semantic search engine based on thousands of Wikipedia articles related to the ocean. Due to the tedious and messy nature of data processing, only the end results (processed text and indexes) has been included in the codebase.
- The ability to generate conversational answers using OpenAI's API (text-davinci-003). Although GPT3.5 has been found to produce the best results, other text summarization and completion LLMs can be used in its place.



- Annotation of generated answers at the sentence level, linking pieces of the answer to their source materials.

## Future developments

- Fact-checking of generated answers, which is a technically challenging problem that even large companies like Bing and Google are still working on. In the meantime, users will have to fact-check the answers themselves by referring to the annotated sources.
- Annotation at the word/phrase level (possibly using non-maximum suppression, inspired by the computer vision bounding-box drawing task).
- Scaling the search engine beyond limited pre-processed data focusing only on the ocean and whales. A possible next step is to develop a general-purpose search engine.
- Improving response speed by finding ways to reduce the time it takes to rely on OpenAI's API.

## Limitations

As previously noted, Cetasearch currently faces several limitations, including speed, the amount of indexed web documents, and the lack of a fact-checking module. These limitations can impact the accuracy and usefulness of search results, and efforts are underway to address these issues and enhance the functionality of the tool.

## Installation

---

1. Make sure you include your OpenAI API key as an environment variable under `OPENAI_API_KEY` name. You can sign up for a 3-month free trial [here](#) and access your API key [here](#)
2. Install requirements `pip install -r requirements.txt`
3. Run `python main.py` to initiate the web server
4. Access the search UI using your web browser at `127.0.0.1:5000`

## Credits & References

---

- [Wiki2txt](#) - for text processing of wikipedia archive
- [Sentence transformer](#) - main component of semantic search
- [OpenAI API](#) - main component for text generation
- [Avatar - The way of water](#) and the Whale conservation efforts by multiple organizations are the inspiration of this project.
- Tech stack: Python (Numpy, Pandas, Flask, etc.), Transformer-based Large Language Models

*Note: As of Feb 19, 2023, Microsoft's New Bing is still in beta testing and not yet release to the public.*