

# Building a Corpus

## Building My Knowledge Base

My knowledge base is in the form of a python dictionary. The dictionary has keys that are some important terms from the scraped corpus, and values that are lists of sentences from the corpus that contain the key term. The knowledge base was created by scraping 20 random websites that were found by crawling the hyperlinks from the Wikipedia page for Frogs. After scraping and processing these sites to extract and count their unique tokens, the 40 most common terms are displayed (shown below).

```
['frog', 'frogs', 'unknown', 'years', 'acquired', 'months', 'bred', 'gender', 'adult', 'wild', 'date', 'species', 'male', 'anura', 'toad', 'female', 'last', 'doi', 'retrieved', 'may', 'id', 'died', 'mb', 'bufo', 'tree', 'deceased', 'rana', 'water', 'eggs', 'reported', 'amphibians', 'museum', 'time', 'skin', 'university', 'use', 'page', 'tadpoles', 'california', 'toads']
```

From these 40 terms, I manually selected 10: ['frog', 'species', 'tree', 'water', 'eggs', 'amphibians', 'tadpoles', 'toad', 'anura', 'skin']

These 10 terms were then made the keys for the lists of sentences that contain them. Shown below is part of the list for the key term 'skin':

```
['The membranes between their toes and loose skin flaps on their sides catch the air as they fall, helping them to glide, sometimes 50 feet or more, to a neighboring tree branch or even all the way to the ground.', "speculation is that iron pigments in the frog's skin interfere with the MRI."), 'When breeding, adult males develop rough pads on their feet called Nuptial Pads which help them to grab and hang on to the slippery skin of a female during amplexus.', 'any frog of the widespread family Ranidae, most members of which are semiaquatic and have smooth, moist skin and relatively long hind legs used for leaping.', 'Their skin varies in colour from well- camouflaged dappled brown, grey and green to vivid patterns of bright red or yellow and black to show toxicity and ward off predators .', 'Frog skin has a rich microbiome which is important to their health.', 'The skin is semi-permeable, making them susceptible to dehydration, so they either live in moist places or have special adaptations to deal with dry habitats.', 'The European fire-bellied toad ( Bombina bombina ) has a slightly warty skin and prefers a watery habitat [3] whereas the Panamanian golden frog ( Atelopus zetekii ) is in the toad family Bufonidae and has a smooth skin.', 'A frog's skin is protective, has a respiratory function, can absorb water, and helps control body temperature.', 'The secretion is often sticky and helps keep the skin moist, protects against the entry of moulds and bacteria, and make the animal slippery and more able to escape from predators.', '[53] The skin is shed every few weeks.', 'The sloughed skin is then worked towards the head where it is quickly eaten.', 'To warm up, they can move into the sun or onto a warm surface; if they overheat, they can move into the shade or adopt a stance that exposes the minimum area of skin to the air.', "[55] The colour of a frog's skin is used for thermoregulation.", "Many frogs are able to absorb water and oxygen directly through the skin, especially around the pelvic area, but the permeability of a frog's skin can also result in water loss.", 'Glands located all over the body exude mucus which helps keep the skin moist and reduces evaporation.', 'Some arboreal frogs reduce water loss by having a waterproof layer of skin, and several South American species coat their skin with a waxy secretion.', 'This reduces the amount of skin exposed to the air or a dry surface, and thus reduces water loss.', 'They contain blood vessels and are thought to increase the area of the skin available for respiration.', '[59] In certain other species, the skin at the top of the head is compacted and the connective tissue of the dermis is co-ossified with the bones of the skull ( exostosis ).', 'Features such as warts and skin folds are usually on ground-dwelling frogs, for whom smooth skin would not provide such effective camouflage.', '[62] Some are even able to control their skin texture.', 'Wood Frog : ( Lithobates sylvaticus or Rana sylvatica ) uses disruptive coloration including black eye markings similar to voids between leaves, bands of the dorsal skin (dorsolateral dermal plica) similar to a leaf midrib as well as stains, spots & leg stripes similar to fallen leaf features.', 'The Titicaca water frog ( Telmatobius culeus ) is one such species and has wrinkly skin that increases its surface area to enhance gas exchange.', 'They do not have external ears ; the eardrums ( ty
```

## Sample Chatbot Dialog

Bot: Hello! I am FrogBot, what is your name?

User: Rahul

Bot: Hello Rahul, what are you interested in?

User: I love toads. What is your favorite amphibian?

Bot: Frogs are my favorite.