Crystal Ngo and Kiara Madeam
Webcrawler Assignment
In a doc: (1) describe how you created your knowledge base, include screen shots of the knowledge base, and indicate your top 10 terms; (2) write up a sample dialog you would like to create with a chatbot based on your knowledge base
Topic: How shrimp (and other animals) absorb visual pigments
Starting Website: https://link.springer.com/article/10.1007/s00359-015-1063-y

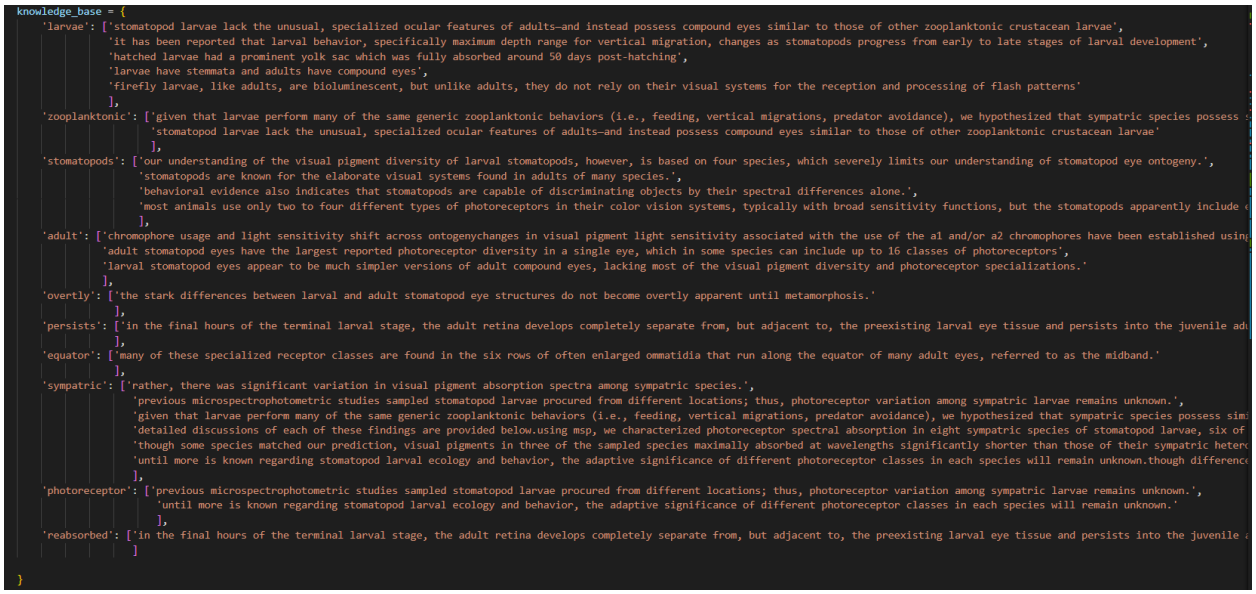
(1) Our top 10 terms are: larvae, zooplanktonic, stomatopods, adult, overtly, persists, equator, sympatric, photoreceptor, reabsorbed

At first, it was difficult to find a link that was interesting and gave us at least 15 relevant websites, but we eventually settled on marine biology and were able to find some research sites that allowed us to scrape their scholarly sources on shrimp and their visual processors. Even after generating a substantial list of relevant urls, we still had to filter out certain links that generated an error as our program attempted to scrape them. For the sites that worked properly, we wrote our scraped contents to a file.

Despite being able to surpass those hurdles, we still had to filter out files that did not give enough relevant information besides the authors and references of the article or information about the website such as their policies. Only after filtering out irrelevant files did we process the text using NLTK and regular expressions so that we could glean relevant terms from the corpora with tf-idf. This equation showed us potential contenders for us to manually choose 10 terms from.

After we handpicked our top 10 terms, we decided to find all the sentences in the corpora that contained those terms and output them to the terminal. There were too many sentences per word for us to see all 10 words' sentences at once, so we limited the search to five sentences each with our most important words. From there, we further handpicked the most interesting and relevant sentences based on our domain knowledge and used those as our relevant knowledge base facts.

We decided to go with a python dictionary for our knowledge base because we felt that this was the easiest format for us to conceptualize, but in the future (for the chatbot assignment) we will look into other knowledge base formats.



```
knowledge_base = {
    'larvae': ['stomatopod larvae lack the unusual, specialized ocular features of adults—and instead possess compound eyes similar to those of other zooplanktonic crustacean larvae',
               'it has been reported that larval behavior, specifically maximum depth range for vertical migration, changes as stomatopods progress from early to late stages of larval development',
               'hatched larvae had a prominent yolk sac which was fully absorbed around 50 days post-hatching',
               'larvae have stemmata and adults have compound eyes',
               'firefly larvae, like adults, are bioluminescent, but unlike adults, they do not rely on their visual systems for the reception and processing of flash patterns'
              ],
    'zooplanktonic': ['given that larvae perform many of the same generic zooplanktonic behaviors (i.e., feeding, vertical migrations, predator avoidance), we hypothesized that sympatric species possess
                      'stomatopod larvae lack the unusual, specialized ocular features of adults—and instead possess compound eyes similar to those of other zooplanktonic crustacean larvae'
                     ],
    'stomatopods': ['our understanding of the visual pigment diversity of larval stomatopods, however, is based on four species, which severely limits our understanding of stomatopod eye ontogeny.',
                    'stomatopods are known for the elaborate visual systems found in adults of many species.',
                    'behavioral evidence also indicates that stomatopods are capable of discriminating objects by their spectral differences alone.',
                    'most animals use only two to four different types of photoreceptors in their color vision systems, typically with broad sensitivity functions, but the stomatopods apparently include
                    ],
    'adult': ['chromophore usage and light sensitivity shift across ontogenychanges in visual pigment light sensitivity associated with the use of the a1 and/or a2 chromophores have been established using
              'adult stomatopod eyes have the largest reported photoreceptor diversity in a single eye, which in some species can include up to 16 classes of photoreceptors',
              'larval stomatopod eyes appear to be much simpler versions of adult compound eyes, lacking most of the visual pigment diversity and photoreceptor specializations.'
             ],
    'overtly': ['the stark differences between larval and adult stomatopod eye structures do not become overtly apparent until metamorphosis.'
               ],
    'persists': ['in the final hours of the terminal larval stage, the adult retina develops completely separate from, but adjacent to, the preexisting larval eye tissue and persists into the juvenile ad
                ],
    'equator': ['many of these specialized receptor classes are found in the six rows of often enlarged ommatidia that run along the equator of many adult eyes, referred to as the midband.'
               ],
    'sympatric': ['rather, there was significant variation in visual pigment absorption spectra among sympatric species.',
                  'previous microspectrophotometric studies sampled stomatopod larvae procured from different locations; thus, photoreceptor variation among sympatric larvae remains unknown.',
                  'given that larvae perform many of the same generic zooplanktonic behaviors (i.e., feeding, vertical migrations, predator avoidance), we hypothesized that sympatric species possess sim
                  'detailed discussions of each of these findings are provided below.using msp, we characterized photoreceptor spectral absorption in eight sympatric species of stomatopod larvae, six of
                  'though some species matched our prediction, visual pigments in three of the sampled species maximally absorbed at wavelengths significantly shorter than those of their sympatric heter
                  'until more is known regarding stomatopod larval ecology and behavior, the adaptive significance of different photoreceptor classes in each species will remain unknown.though differenc
                  ],
    'photoreceptor': ['previous microspectrophotometric studies sampled stomatopod larvae procured from different locations; thus, photoreceptor variation among sympatric larvae remains unknown.',
                      'until more is known regarding stomatopod larval ecology and behavior, the adaptive significance of different photoreceptor classes in each species will remain unknown.'
                     ],
    'reabsorbed': ['in the final hours of the terminal larval stage, the adult retina develops completely separate from, but adjacent to, the preexisting larval eye tissue and persists into the juvenile
                  ]

}
```

(2)
Hello, I'm Shrimpy, a chatbot specializing in how marine animals see. What's your name?

Hi Shrimpy! I'm Crystal.

Hi Crystal! Do you like marine biology?

Yeah! Why are you named Shrimpy?

Why are you named Crystal? I'm named after the mantis shrimp, or stomatopod, which has the most complex color vision of any animal!

What does that mean?

That means stomatopods can see colors that people like us can't see. They can also see UV and polarized light.

That's really cool! Wait, are you human?

Moving on!