Niraj Amalkanti, Bennett Ng

10/10/2013

BioE 131/231

Project 1

Our code for this project interacts with the NCBI databases to fetch data about selected organisms and generate appropriate cards to represent them. The code consists of two main components:

* The Scraper utilizes BioPython to query NCBI and fetch data from the “genome” and “nucleotide” databases. Furthermore, a Google Image Search is performed to obtain URLs to pictures of each specified organism. These data are assembled into an Entry object which represents one organism.
  + The scraper supports two access methods – one simple one to fetch a specified number of Entries, and one to fetch a specific list of entries (by accession number).
* The HTMLGenerator accepts Entry objects and generates and HTML document. The HTML code is generated per XHTML specifications and is essentially a mass of div objects, each div corresponding to a card. The CSS file styles the cards so they can more effeciently be printed out. The only interface is that a list of entries is passed into the constructor and get\_html method returns the HTML as an etree node.We used the lxml library for our XHTML generation. Our CSS file was specified manually(though is reused).
  + The HTMLGeneratorTest is the test class for the HTMLGenerator. We used the default Python unitttest framework to build it.
* The Entry object specified the interface used between the scraper and the HTML generator. We implemented the builder pattern to create the object.

Bennett wrote the Scraper and Niraj wrote the HTMLGenerator. The Entry object was a collaborative effort.