1. Basics

**References:**

1. Speech and Language Processing, by Dan Jurafsky and James H. Martin. Prentice Hall Series in Artificial Intelligence, 2008.
2. Natural Language Processing with Python, by Steven Bird, Ewan Klein and Edward Loper, 2014.

Quick Review

Postman is a powerful tool for performing integration testing with your API. It allows for repeatable, reliable tests that can be automated and used in a variety of environments and includes useful tools for persisting data and simulating how a user might actually be interacting with the system.

Lab Practices

* 1. Requests and responses

Postman is available as a native app for macOS, Windows, and Linux operating systems. To install Postman, go to the apps page available on <https://www.getpostman.com> and click Download for macOS / Windows / Linux depending on your platform.

* 1. Universal Part-of-Speech Tagset

Tagged corpora use many different conventions for tagging words. To help us get started, we will be looking at a simplified tagset.

* **The Lookup Tagger**

A lot of high-frequency words do not have the NN tag. Let's find the hundred most frequent words and store their most likely tag. We can then use this information as the model for a "lookup tagger" (an NLTK UnigramTagger):

Put all this together:

Exercises

1. Search the web for "spoof newspaper headlines", to find such gems as: British Left Waffles on Falkland Islands, and Juvenile Court to Try Shooting Defendant. Manually tag these headlines to see if knowledge of the part-of-speech tags removes the ambiguity.
2. Tokenize and tag the following sentence: They wind back the clock, while we chase after the wind. What different pronunciations and parts of speech are involved?
3. Write programs to process the Brown Corpus and find answers to the following questions:
   1. Which nouns are more common in their plural form, rather than their singular form? (Only consider regular plurals, formed with the -s suffix.)
   2. Which word has the greatest number of distinct tags? What are they, and what do they represent?
   3. List tags in order of decreasing frequency. What do the 20 most frequent tags represent?
   4. Which tags are nouns most commonly found after? What do these tags represent?